UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Copula Models for Dependence: Comparing Classical and Bayesian Approaches

Lídia Maria Branco Correia Martins André

Mestrado em Estatística e Investigação Operacional

Especialização em Estatística

Dissertação orientada por:

Professora Doutora Patrícia de Zea Bermudez

2019

Acknowledgements

First and foremost I wish to express my sincere gratitude to my advisor, Professor Patrícia de Zea Bermudez, who initiated me into the subject of Copula modelling, for her guidance, patience, support and helpful advice during the past months.

I wish to thank to Sofia Antunes, Inês Cardoso, Andreia Silva and Inês Diogo for their friendship, encouragement and continuous support throughout these months. I am especially grateful to my friend Hugo Moreira for his untiring help whenever I needed advice on statistical matters and for sharing his ideas, which were extremely helpful to improve this work.

A special word of gratitude is due to my family for all the love they gave me throughout the years, for their encouragement and for allowing me to pursue my studies. I would also like to thank for their valuable help in the writing of this dissertation, for critical and helpful comments on early drafts and mostly with LaTeX language.

I would like to thank Professor Paula Pereira for supplying the script for the Portugal map and for all the help in adapting it.

I would also like to thank Instituto Português do Mar e da Atmosfera (IPMA) for allowing the use of the observed wind speed data and Instituto Dom Luiz, Universidade de Lisboa, for providing the simulated data.

Finally I would also like to extend thanks to those who, in a way or another, helped and encouraged me to write this thesis.

This dissertation was elaborated under the project with title “Data Fusion and Calibration Methods for Spatial Risk Analysis” with reference: PTDC/MATSTA/28649/2017, supported by FCT, Fundação para a Ciência e Tecnologia.

Resumo

O coeficiente de correlação de Pearson é usado para quantificar a intensidade de associações lineares entre duas variáveis. Alternativamente, coeficientes de correlação não paramétricos, como os de Spearman e Kendall, podem também ser utilizados. Contudo, em aplicações reais, as associações raramente são lineares. Nestas situações, é interessante ajustar um modelo multivariado às variáveis aleatórias que descreva outros tipos de relações de dependência. Embora se possa automaticamente pensar no modelo Normal tradicional multivariado, este geralmente não consegue descrever apropriadamente a complexidade das associações existentes entre as variáveis. Além disso, dificilmente consegue modelar adequadamente dados que apresentem, por exemplo, caudas pesadas ou assimetrias acentuadas. As cópulas ultrapassam este problema pois permitem descrever a dependência conjunta entre várias variáveis. Numa estrutura multivariada, é possível criar a distribuição conjunta das diversas variáveis aleatórias independentemente das suas distribuições marginais. As cópulas têm sido amplamente usadas em diversas áreas, principalmente na área da banca e dos seguros. Contudo, devido à sua versatilidade, também têm tido grande aplicação em questões ambientais e climáticas.

Nesta dissertação, são utilizadas cópulas para analisar a dependência entre duas variáveis: a velocidade máxima diária de vento, X, medida em km/h, observada em 40 estações meteorológicas localizadas em Portugal continental desde 2000 até 2012 e a velocidade máxima diária de vento simulada, Y, produzida por um simulador com uma grelha regular com células de 81 km^2. Um dos principais benefícios em usar os dados simulados face aos observados é não haver presença de valores em falta. Em algumas estações a proporção de valores omissos (NA) chega a atingir os 90%. Por esse motivo, das 117 estações meteorológicas do Instituto Português do Mar e da Atmosfera localizadas no continente, só foram consideradas as que apresentavam menos de 30% de NAs, i.e., 40 estações. Os dados observados e simulados irão ser analisados por estação do ano e, em todas as estações do ano, só uma observação em cada cinco será considerada, de forma a minimizar a dependência de curto prazo existente em cada série. Foram também retiradas do estudo as velocidades de vento iguais a 0 por serem, possivelmente, erros da torre de medição ou valores em falta.

O maior problema de usar os dados simulados face às velocidades de vento diárias registadas prende-se com o facto de, nalgumas estações, embora possa haver uma boa correspondência no centro da distribuição, as caudas tendem a ser bastante diferentes, especialmente no que se refere à cauda superior. Os dados simulados apresentam tipicamente caudas direitas menos pesadas do que as dos dados observados. Outro problema que pode surgir ao serem utilizados dados simulados relaciona-se com a localização destes. Nalgumas situações, maioritariamente na Primavera e, ocasionalmente no Verão, as velocidades de vento simuladas parecem ter sofrido um deslocamento para a direita face às velocidades de vento registadas. Num contexto ambiental, valores extremos de velocidade de vento podem causar vários danos materiais, nomeadamente no que diz respeito a redes eléctricas, infra-estruturas agrícolas, fabris e/ou serviço público, ou danos na via pública. Na ocorrência de ventos fortes, se uma rede de energia ficar danificada, e, consequentemente, uma localidade ficar sem energia durante um certo período de tempo, cabe à empresa de energia reembolsar a população. No caso de infra-estruturas agrícolas ou fabris, que estão em constante produção, um corte de energia, provocado por eventos extremos de ventos, implica a paragem da produção. Em ambos os casos, quanto maior for o período sem electricidade, maior será o prejuízo. Por outro lado, o conhecimento do comportamento do vento pode ser importante para questões municipais, nomeadamente numa planificação urbanística adequada à zona. Se esta for demasiado ven-

tosa, não deverão ser construídos terraços ou plantadas árvores que, na ocorrência de eventos extremos, poderão cair e danificar a via pública ou até habitações/estabelecimentos. Deste modo, estudar e entender a dependência entre as velocidades de vento diárias registadas e os valores simulados é extremamente importante.

Apesar das cópulas permitirem a separação da modelação das distribuições univariadas e da estrutura de dependência conjunta, modelar adequadamente as velocidades máximas diárias de vento registadas e as velocidades simuladas produzidas pelo simulador é importante. Foram consideradas 4 distribuições para modelar as velocidades do vento: a Lognormal, a Gama, a Weibull e a Burr com 3 parâmetros e foram realizados vários testes de ajustamento, tais como o Qui-Quadrado ou o Kolmogorov-Smirnov. Os resultados mostraram que a Lognormal e a Gama são as distribuições que melhor se ajustam aos dados de vento e, contrariamente ao que seria de esperar com base na literatura, a Weibull parece ser a que menos vezes se ajusta; ver [Mert and Karakus, 2015] ou [Shepherd, 1978]. Para modelar a dependência conjunta, foram consideradas cinco famílias de cópulas: a Gaussiana, a Student *t*, a Clayton, a Frank e a Gumbel. Ocasionalmente, foi ajustada a cópula Joe. Foram ainda usados três testes semi-paramétricos de ajustamento para cópulas baseados no conceito de cópula empírica, da transformação de Kendall e da transformação de Rosenblatt. A cópula Gumbel foi ajustada 72 vezes em 160 (4 estações do ano $\times$ 40 estações meteorológicas), o que constitui 45% dos casos. Esta cópula é caracterizada por apresentar dependência na cauda superior, o que significa que existe dependência para valores altos da velocidade de vento registada e simulada. No global, 65% das cópulas ajustadas apresentam dependência na cauda superior, enquanto cerca de 32% não apresenta dependência nas caudas, pelo que os ventos simulados e observados se comportam similarmente ao longo do suporte.

Foram também discutidos e comparados diferentes tipos de estimação do parâmetro da cópula. Estes estão englobados em 3 categorias: estimação paramétrica, estimação semi-paramétrica e estimação não paramétrica. No primeiro caso, existem 2 métodos, o método da máxima verosimilhança e o método *"Inference for Margins"*, onde são consideradas as distribuições marginais de cada variável. No caso da estimação pelo método da máxima verosimilhança, a estimação dos parâmetros das distribuições univariadas e do parâmetro da cópula são obtidas conjuntamente, enquanto o segundo método está dividido em 2 fases: na primeira, os parâmetros das cópulas são estimados e, na segunda, os parâmetros obtidos na primeira fase são utilizados para estimar o parâmetro da cópula. Por outro lado, a estimação semi-paramétrica engloba o método da máxima pseudo-verosimilhança e tem como base as distribuições empíricas das variáveis. Por último, a cópula pode ser estimada não parametricamente pela cópula empírica ou pelas medidas de dependência não paramétricas, tais como o τ de Kendall e o ρ de Spearman, que têm uma relação directa com as cópulas.

Depois de ajustados e estimados os parâmetros, foram simuladas observações a partir da cópula e comparadas com as velocidades de ventos reais e com as obtidas pelo simulador. Foram também apresentadas estimativas da dependência conjunta de ocorrência de ventos fortes. Observou-se que a dependência entre as variáveis é superior nas estações do Outono e do Inverno e menor no Verão, o que seria de esperar.

Por fim, aplicou-se a abordagem Bayesiana a 9 estações com o intuito de a comparar com a abordagem clássica. No geral, não se obtiveram estimativas muito diferentes das que foram obtidas anteriormente. Contudo, a amplitude dos intervalos de credibilidade a 95% revelou-se ser inferior à amplitude dos intervalos de confiança a 95%, o que constitui uma vantagem em utilizar a abordagem Bayesiana face à abordagem clássica.

Palavras-Chave: Dependência, Cópulas, Velocidade de Ventos, Estatística Bayesiana

Abstract

Pearson's correlation coefficient is commonly used for quantifying the intensity of linear associations between two variables. Alternatively, non-parametric correlation coefficients, such as Spearman's and Kendall's can also be applied. However, in real data applications, these associations are often nonlinear. In these situations it would be of great interest to fit a multivariate model to the random variables which may describe other types of dependence structures. Although one may automatically think about the traditional Gaussian multivariate model, this model does not generally describe the complexity of the association between the variables. Moreover, it can hardly be adequate to model data which show, e.g., strong asymmetries or heavy tails. Copulas overcome this drawback. They allow the description of the joint dependence of several variables. In a multivariate framework, they enable to create the joint distribution of the vector of random variables independently of their marginal distributions. In this thesis, a copula approach is used to analyse the dependence between daily maximum wind speeds, X, (km/h) observed in 40 stations spread out in the continental part of Portugal from 2000 to 2012 and simulated daily maximum wind speeds, Y, produced by a simulator at a regular grid of 81 km^2 grid cell size. One of the major benefits of using the simulated data is that it has no missing values while the observed data has an extremely high proportion of missing observations, for instance in some stations it reaches 90%. The main problem is that the simulated and the observed daily maximum wind speeds, in some stations, do not match well and tend to differ, mostly in the right tail. Consequently, it is very important to understand the dependence between X and Y. Four distributions were considered to model the wind speed data: the Lognormal, the Gamma, the Weibull and the 3-parameter Burr distributions and five family of copulas were considered to model (X,Y): Gaussian, Student *t*, Clayton, Frank and Gumbel. Occasionally, Joe's copula was fitted. The results showed that Lognormal and Gamma are the most suitable marginal distributions and Gumbel's Copula is the most adequate to model the dependence. Finally, the classical modelling is compared with a Bayesian approach. The Bayesian estimates for the parameters of the univariate distributions and for the copula proved to be not very different. However, it provided narrower credible intervals for the estimates, which constitutes an advantage towards the classical approach.

Keywords: Dependence, Copulas, Wind Speed, Bayesian Statistics

Contents

List of Figures

List of Tables

List of Acronyms

MLE maximum likelihood estimation. 21

IFM Inference Functions for Margins. 22

MPLE maximum pseudo-likelihood estimator. 22

KS Kolmogorov-Smirnov. 25

CvM Cramér-von-Mises. 25

AD Anderson-Darling. 25

AIC Akaike's Information Criterion. 25

BIC Bayesian Information Criterion. 25

DIC Deviance Information Criterion. 25

S-W Shapiro-Wilk. 25

Notation

I Unit segment. 3

$C(u,v)$ Bivariate Copula. 3

$\overline{C}(u,v)$ Survival Copula. 4

ρ Pearson's correlation coefficient. 6

τ Kendall's tau. 7

ρ_S Spearman's rho. 7

λ_U upper tail dependence. 9

λ_L lower tail dependence. 9

$\varphi^{[-1]}$ pseudo-inverse. 15

C_n empirical copula. 23

χ^2 Chi-Square. 25

R^2 coefficient of determination. 25

Π Product Copula. 49

C_ρ^{gaus} Gaussian Copula. 49

$C_{\rho\eta}^t$ Student t-Copula. 49

C_α^c Clayton Copula. 49

C_α^{gu} Gumbel Copula. 49

C_α^f Frank Copula. 49

C_α^j Joe Copula. 49

C_α^{sc} Survival Clayton Copula. 49

C_α^{sg} Survival Gumbel Copula. 49

1 | Introduction

Due to the randomness of the wind speed, estimating its behaviour is hard, therefore predicting future wind events accurately is almost impossible. In order to simplify the prediction of the wind speed behaviour and avoid possible damages when an extreme event occurs, simulating wind speeds from the real daily maximum winds speeds may be of great importance. Moreover, if there is an exact match between the two variables, one can rely on the values produced by the simulator instead of considering the real ones which may have missing values. In fact, in 117 meteorological stations spread out in the continental part of Portugal, just 40 stations show a proportion of missing values lower than 30%.

However, a problem which arises from the use of simulated data instead of observed data is the fact that, in some cases, they do not match nicely, especially for extreme values. In particular, the right tails tend to differ from each other. For the simulated data, they tend to be less heavier than for the observed data. Furthermore, the mapping of the risk based on simulated data tends to underestimate the damage probability. In an environmental context, extreme wind speeds might cause damage to property, for instance regarding to power networks, agricultural or factory infrastructures, public service institutions, such as hospitals, or damage to public roads, for example tree fallings, erosion or landslides. For instance, strong winds can cause power energy loss in a region which may affect its infrastructures. Moreover, for longer periods without energy, the damage will certainly be larger. On the other hand, the knowledge of the wind behaviour can be extremely useful for municipal issues, for instance in urban planning appropriate to the area. If it is an extremely windy zone, no terraces should be built or trees planted, since they can fall and damage surrounding houses. Regarding to public service, such as hospitals, even though they may be self-sufficient, long periods without power energy caused by extreme occurrences of wind may prevent medical procedures from being performed. For these reasons, the study of the dependence between the two type of data is of great relevance, so that improvements on the observed wind speed data can be made based on the simulated data.

But why is the use of Copulas important? Indeed, for measuring and quantifying the linear associations between variables, we can use, not only Pearson's correlation coefficient, but also non-parametric coefficients, such as Kendall's tau and Spearman's rho. However, these coefficients only provide information about the overall strength of dependence and do not reveal how it varies across the distribution; see [Kostova *et al*, 2012]. Moreover, in real applications, the associations between variables are often non-linear. Furthermore, fitting a multivariate model to random variables can be useful to measure other types of dependence. Then we can think about the tradicional Gaussian multivariate model. Nonetheless, it is hardly adequate since generally it does not describe complex associations between variables, such as heavy tails or strong asymmetries.

In recent years, there has been a growing interest in copulas and their applications. They first appeared in [Sklar, 1959] and the theory about copulas is presented in detail in [Nelsen, 2006] and [Joe, 1997]. They have been widely applied in insurance and finance problems; see [Dana, 2007], [Embrechts *et al*, 2003], [Shemyakin and Youn, 2006], [Erntell, 2013] or [Dos Santos, 2011]. However, copulas start to be applied to other fields, such as climate; see for instance [Cong and Brady, 2012], [Diaz, 2017] or [Kostova *et al*, 2012].

Copulas enable to create the joint distribution of the vector of random variables independently from their marginal distributions. In a bivariate framework, according to Sklar's theorem, if X and Y are two random variables with joint distribution $H(x,y) = P(X \leq x,\ Y \leq y)$ and distribution functions

$F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$, then there exists a copula C such that

$$H(x,y) = C(F(x), G(y)). \tag{1.1}$$

Moreover, if X and Y are continuous variables, C is unique. Furthermore, if $u = F(x)$ and $v = G(y)$ are distribution functions, $C(u,v) = C(F(x), G(y))$ is a valid bivariate distribution. Finally, copulas provide a more flexible methodology for modelling the dependence between several variables, indicating in which part of the distribution the dependence is stronger; see [Kostova *et al*, 2012].

The aim of this dissertation is to model the dependence between daily maximum wind speeds, measured in km/h, observed in 40 stations spread out in the continental part of Portugal from 2000 to 2012, and simulated wind speeds produced by a simulator at a regular grid of 81 km^2 grid cell size. The end purpose, which is beyond the scope of this thesis, is to use the simulated wind speeds after being calibrated using the observed data, i.e., to bring the simulated wind speeds in line with the observed wind speeds. For accomplishing this, it is very important to understand and characterise the dependence existing between the simulated and the observed data.

This dissertation will be organised as follows. In **Chapter 2**, a theoretical background on dependence modelling with copulas is given and the most common copula families are presented. In **Chapter 3**, we present the statistics of copulas. Several methods for estimating the copula association parameter are enumerated and the Bayesian inference for copulas is approached. We then assess the fit of the model. First, we address the fit of marginal distributions and then the fit of copulas to the data sets. Although copulas allow to model the dependence independently from the marginal distributions, representing the wind speed data by an adequate probability function is important. The Weibull distribution appears in the literature as the most common distribution to model wind type data; see [Harris and Cook, 2014], [Mert and Karakus, 2015], [Pobocikova *et al*, 2017] or [Shepherd, 1978]. Nevertheless, we consider three additional distributions: the Lognormal, the Gamma and the 3-parameter Burr; see [Mert and Karakus, 2015] or [Shepherd, 1978]. The fact that all distributions are unspecified, that is, all parameters have to be estimated from the data, is also addressed in regard to the performance of the usual goodness-of-fit tests. For copulas, we enumerate several possible goodness-of-fit tests, based on the results of [Genest *et al*, 2009], and some other criteria which are useful to select the appropriate copula.

In **Chapter 4**, we present the procedure used to model the winds in the region of Castelo Branco. A full analysis is made for the Winter season and the results are shown for nine meteorological stations spread out in Portugal. The results for the other 31 stations are presented in the Appendices due to space constraints. The Bayesian approach, as well as a comparison between the copula estimation methods, are made just for the nine selected meteorological stations.

2 | Copula Theory

2.1 Introduction

In this chapter, we summarise some well-known definitions and standard results about copulas. We will use the notation employed by [Shemyakin and Kniazev, 2017]: $I = [0,1]$ is the unit segment, $I^2 = [0,1] \times [0,1]$ is the unit square, and, for any $0 \leq u_1 \leq u_2 \leq 1$, $0 \leq v_1 \leq v_2 \leq 1$, $B = [u_1, u_2] \times [v_1, v_2]$ is a rectangular region in the unit square. We will also restrict ourselves to the bivariate case. Extensions to the multivariate case can be found in the literature, such as [Joe, 2014] or [Embrechts *et al*, 2003]. The structure of the chapter is mainly based on [Embrechts *et al*, 2003] and [Shemyakin and Kniazev, 2017].

Definition 2.1.1. Let $A(u,v)$ be a function from I^2 to I and B be a rectangular region of the unit square.

1. The A**-volume** of the region B is given by:

$$V_A(B) = A(u_2, v_2) - A(u_1, v_2) - A(u_2, v_1) + A(u_1, v_1). \tag{2.1}$$

2. $A(u,v)$ is **quasi-monotone** if, for any rectangular area B in the unit square, its A-volume is nonnegative.

3. $A(u,v)$ is **grounded** on I^2 if $A(0,v) = A(u,0) = 0$ for any $u, v \in I$.

4. $A(u,v)$ is 2**-increasing** if $V_A(B) \geq 0$, for any rectangular area B in the unit square.

Note that any grounded nonnegative quasi-monotone function on I^2 is increasing in each argument.

Definition 2.1.2. A 2**-dimensional copula** is a function C with domain I^2 satisfying the following properties:

1. For any u, v in I, $C(0,v) = C(u,0) = 0$.

2. For any u, v in I, $C(1,v) = v$, $C(u,1) = u$.

3. For any rectangular region $B \subseteq I^2$, $\quad V_C(B) \geq 0$.

The partial derivatives, $\frac{\partial C}{\partial u}$ and $\frac{\partial C}{\partial v}$, of any copula $C(u,v)$ exist for almost all $u, v \in I$. Suppose that $\frac{\partial^2 C}{\partial u \partial v}$ and $\frac{\partial^2 C}{\partial v \partial u}$ exist and are continuous in I^2. The *copula density* is then given by

$$c(u,v) = \frac{\partial^2 C}{\partial u \partial v} = \frac{\partial^2 C}{\partial v \partial u}. \tag{2.2}$$

It follows that, if $u = F(x)$ and $v = G(y)$ are distribution functions, then any copula of the form $C(u,v) = C(F(x), G(y))$ is a valid bivariate distribution function, and the converse is also true. These statements can be found in [Sklar, 1959].

2.2 Sklar's Theorem

Sklar's Theorem first appeared in 1959 and is central to the theory of copulas. It states that any multivariate distribution can be represented as the composition of a copula and its marginal distributions.

Theorem 2.2.1. *Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all x, y*

$$H(x,y) = C(F(x), G(y)). \tag{2.3}$$

If F and G are continuous, then C is unique. Otherwise, C is uniquely determined on $\mathrm{Ran}(F) \times \mathrm{Ran}(G)$*, where* $\mathrm{Ran}(F)$ *and* $\mathrm{Ran}(G)$ *are the ranges of F and G, respectively. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by* (2.3) *is a joint distribution function with margins F and G.*

For the proof, see [Nelsen, 2006].

Sklar's theorem states, not only that every copula with marginal distributions as arguments is a valid bivariate distribution, but also that is possible to represent every valid bivariate distribution as a copula of its marginals. For the continuous case, it always allows to separate the univariate marginals from the dependence structure. In the case of discrete one-dimensional marginal distributions, one cannot assume that (2.3) is unique.

Corollary 2.2.1.1. *Let H be a joint distribution function with margins F and G, C a copula defined by* (2.3)*,* $F^{(-1)}(u) = \inf\{x \in \mathbb{R} \mid F(x) \geq u\}$ *and* $G^{(-1)}(u) = \inf\{y \in \mathbb{R} \mid G(y) \geq u\}$*. Then, for any* $u, v \in I$

$$C(u,v) = H(F^{(-1)}(u), G^{(-1)}(v)). \tag{2.4}$$

The inversion method for constructing copulas for joint distribution functions is provided by (2.4), when F and G are continuous. See [Nelsen, 2006] for the case when F and G are discrete marginal distributions.

One property of copulas is that they are either invariant or do not change much for strictly monotone transformations of random variables, in the case where F and G are continuous marginal distributions; see [Embrechts *et al*, 2003] or [Nelsen, 2006].

Theorem 2.2.2. *Let X and Y be continuous random variables with copula* C_{XY} *and* $u, v \in I$ *be their marginals,* $u = F(x)$ *and* $v = G(y)$*. Let* α *and* β *be strictly monotone on* $\mathrm{Ran}(X)$ *and* $\mathrm{Ran}(Y)$*, respectively.*

1. *If* α *and* β *are strictly increasing, then*

$$C_{\alpha(X)\beta(Y)}(u,v) = C_{XY}(u,v). \tag{2.5}$$

   In particular, C_{XY} *is invariant under strictly increasing transformations of* X *and* Y*.*

2. *If* α *is strictly increasing and* β *is strictly decreasing, then*

$$C_{\alpha(X)\beta(Y)}(u,v) = u - C_{XY}(u, 1-v). \tag{2.6}$$

3. *If* α *is strictly decreasing and* β *is strictly increasing, then*

$$C_{\alpha(X)\beta(Y)}(u,v) = v - C_{XY}(1-u, v). \tag{2.7}$$

4. *If α and β are strictly decreasing, then*

$$C_{\alpha(X)\beta(Y)}(u,v) = u + v - 1 + C_{XY}(1-u, 1-v). \tag{2.8}$$

This copula is known as the **survival copula**, *it will be denoted by $\overline{C}(u,v)$ and satisfies all copula properties.*

For the proof, see [Nelsen, 2006].

Let X and Y be random variables with distribution functions $F(x)$ and $G(y)$, respectively. The probability that X occurs after time x is called the survival function and it is given by $S_1(x) = P[X > x] = 1 - F(x)$. Similarly, $S_2(y) = 1 - G(y)$ is the survival function for Y. The *joint survival function* is given by $S(x,y) = P[X > x, Y > y]$. [Nelsen, 2006] showed that there is a relationship between univariate and joint survival functions as there is with univariate and joint distribution functions, according to Sklar's theorem. One just has to consider the survival functions as $S_1(x)$ and $S_2(y)$ and then it is straightforward that

$$S(x,y) = \overline{C}(S_1(x), S_2(y)). \tag{2.9}$$

When it is more natural or easier to use survival functions, (2.9) is a useful alternative to $C(u,v)$; see [Shemyakin and Kniazev, 2017].

It is easily noted that X and Y are independent if and only if its copula is

$$\Pi(u,v) = uv, \quad u,v \in I, \tag{2.10}$$

known as *Product Copula*. Additionally, if one variable, say Y, is a function of the other, say X, the copula C must either be the maximum copula, $C(u,v) = M(u,v) = \min(u,v)$, and in this case, Y is monotone increasing in X, or the minimum copula, $C(u,v) = W(u,v) = \max(u+v-1,0)$, and Y is a decreasing function of X; see [Genest and Favre, 2007]. It is easy to show that these three functions satisfy all copula properties.

Theorem 2.2.3. *For any copula C and any $u,v \in I$, the following inequalities hold:*

$$W(u,v) \leq C(u,v) \leq M(u,v). \tag{2.11}$$

The functions $W(u,v)$ and $M(u,v)$ are known as the *lower and upper Fréchet-Hoeffding Bounds*, respectively.

2.3 Dependence Modelling

In order to study the dependence between two random variables, X and Y, and then to build a mathematical model of their dependence, it is required the study of the character and strength of the dependence. The ultimate goal is to fit the data into a functional model of the relationship; see [Shemyakin and Kniazev, 2017]. The fact that many of the properties and measures of dependence remain unchanged under strictly increasing transformations of the random variables is very appealing and, as follows from Theorem 2.2.2, so is the use of a copula to capture the properties of joint distributions which are "scale-invariant"; see [Nelsen, 2006]. The most frequently used measure of dependence is Pearson's linear correlation coefficient, which is a numerical characteristic. However, it is not a copula-based measure of dependence, which can lead to deception and so it should not be used as the "main" measure of dependence. To overcome this issue there exist other measures of dependence such as Kendall's concordance and Spearman's rank correlation. These measures are concordance, and copula-based, measures of dependence; see [Embrechts *et al*, 2003]. There are also graphical ways to assist the modelling of the dependence between two variables, such as the χ-Plots and the K-Plots. The explanation of these graphical methods can be found in detail in [Genest and Favre, 2007].

2.3.1 Linear Correlation

Definition 2.3.1. Let $(X,Y)^T$ be a vector of random variables with nonzero finite variances. The *linear correlation coefficient* for $(X,Y)^T$ is

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}, \tag{2.12}$$

where $\operatorname{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$ is the covariance of $(X,Y)^T$, and $\operatorname{Var}(X) = \mathrm{E}[(X-\mathrm{E}[X])^2]$ and $\operatorname{Var}(Y) = \mathrm{E}[(Y-\mathrm{E}[Y])^2]$ are the variances of X and Y, respectively.

The linear correlation coefficient (2.12) can be estimated by

$$\hat{\rho}(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}, \tag{2.13}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i$ are the sample means of X and Y, respectively, and n is the sample size of a bivariate sample (x_i,y_i), $i = 1,\ldots,n$.
Expression (2.12) is known as Pearson's correlation coefficient ρ and has the following properties:

1. $-1 < \rho(X,Y) < 1$.
2. For a perfect linear correlation, $|\rho(X,Y)| = 1$.
3. If X and Y are independent, then $\rho(X,Y) = 0$. The reciprocal is not true.
4. For $\alpha,\gamma \in \mathbb{R}\setminus\{0\}$, $\beta,\delta \in \mathbb{R}$, $\rho(\alpha X+\beta,\gamma Y+\delta) = \operatorname{sgn}(\alpha\gamma)\rho(X,Y)$, i.e, ρ is invariant under strictly increasing linear transformations [(a)].

Additionally,

- If $\rho(X,Y) > 0$, then X and Y are positively correlated.
- If $\rho(X,Y) < 0$, then X and Y are negatively correlated.
- If $\rho(X,Y) = 0$, then X and Y are linearly independent although not independent. In fact, there might be some other type of relation between X and Y.

Pearson's correlation coefficient can be considered parametric, since it is obtained through basic distribution parameters, such as means and variances; see [Shemyakin and Kniazev, 2017].

Since ρ is easy to estimate and, in the case of elliptical distributions (see Definition 2.4.1), it is a natural scalar measure of dependence, (2.12) is widely used. However, it has some drawbacks. It is not invariant to non linear transformations, for instance $\rho(F(X),G(Y)) \neq \rho(X,Y)$, and thus it can not be defined as a copula function. It is also not robust which means that an observation can have a large influence on its value and therefore it shall not be used when there are, for example, outliers in the sample; see [Embrechts *et al*, 2003].

[(a)] $\operatorname{sgn}(\alpha\gamma)$ is the *sign function*: $\operatorname{sgn}(\alpha\gamma) = 1$ if $\alpha\gamma > 0$ and $\operatorname{sgn}(\alpha\gamma) = -1$ if $\alpha\gamma < 0$

2.3.2 Concordance Measures

To overcome the drawbacks of Pearson's correlation coefficient and to get rid of its parametric structure, one can consider other correlation measures based on ranks such as Kendall's tau and Spearman's rho, which can be considered as non-parametric measures of dependence that measure a type of dependence called concordance.

Definition 2.3.2. Let $(x,y)^T$ and $(\tilde{x},\tilde{y})^T$ be two observations of the vector $(X,Y)^T$ of continuous random variables. Then $(x,y)^T$ and $(\tilde{x},\tilde{y})^T$ are said to be **concordant** if $(x-\tilde{x})(y-\tilde{y}) > 0$ and **discordant** if $(x-\tilde{x})(y-\tilde{y}) < 0$.

Theorem 2.3.1. *Let $(X,Y)^T$ and $(\tilde{X},\tilde{Y})^T$ be independent vectors of continuous random variables with joint distribution functions H and $\tilde{H}$, respectively, with common margins F of X and $\tilde{X}$ and G of Y and $\tilde{Y}$. Let C and $\tilde{C}$ denote the copulas of $(X,Y)^T$ and $(\tilde{X},\tilde{Y})^T$, respectively, so that $H(x,y) = C(F(x),G(y))$ and $\tilde{H}(x,y) = \tilde{C}(F(x),G(y))$. Let Q denote the difference between the probability of concordance and discordance of $(X,Y)^T$ and $(\tilde{X},\tilde{Y})^T$, i.e, let*

$$Q = P[(X-\tilde{X})(Y-\tilde{Y}) > 0] - P[(X-\tilde{X})(Y-\tilde{Y}) < 0]. \tag{2.14}$$

Then

$$Q = Q(C,\tilde{C}) = 4\iint_{I^2} \tilde{C}(u,v)\,\mathrm{d}C(u,v) - 1. \tag{2.15}$$

For the proof, see [Embrechts *et al*, 2003].

Kendall's Tau

As we can see from the following definition, Kendall's tau τ will be simply the probability of concordance minus the probability of discordance; see [Embrechts *et al*, 2003].

Definition 2.3.3. Kendall's tau for a random vector $(X,Y)^T$ is defined as

$$\tau(X,Y) = P[(X-\tilde{X})(Y-\tilde{Y}) > 0] - P[(X-\tilde{X})(Y-\tilde{Y}) < 0], \tag{2.16}$$

where $(\tilde{X},\tilde{Y})^T$ is an independent copy of $(X,Y)^T$.

Denoting by A the number of concordant pairs and by B the number of discordant pairs and, noting that there are $\binom{n}{2}$ distinct pairs (x,y) and $(\tilde{x},\tilde{y})$ in the bivariate sample of size n, Kendall's tau (2.71) can be estimated by

$$\hat{\tau}(X,Y) = \frac{A-B}{A+B} = \binom{n}{2}^{-1}(A-B). \tag{2.17}$$

As stated before, Kendall's tau is a copula-based measure of dependence. Therefore, it is possible to rewrite it as a function of a copula, which will be very useful to estimate the copula parameter in both classical and Bayesian approaches, as we will see later.

Theorem 2.3.2. *Let $(X,Y)^T$ be a vector of continuous random variables with copula C. Kendall's tau for $(X,Y)^T$ is given by*

$$\tau(X,Y) = Q(C,C) = 4\iint_{I^2} C(u,v)\,\mathrm{d}C(u,v) - 1. \tag{2.18}$$

Equation (2.18) is the expected value of the copula with uniform marginals U and V, $C(U,V)$. Therefore,

$$\tau(X,Y) = 4\,\mathrm{E}[C(U,V)] - 1. \tag{2.19}$$

Spearman's Rho

Another concordance measure of dependence, which is also copula-based, is Spearman's rho ρ_S. Although it is also very useful when estimating the copula parameter, one has to be careful with the ties of the sample. This is not a real obstacle when we consider Kendall's tau since, assuming that X and Y are continuous, ties occur with probability 0; see [Genest and Favre, 2007].

Definition 2.3.4. Spearman's rho for the random vector $(X,Y)^T$ is defined as

$$\rho_S(X,Y) = 3(P[(X-X_1)(Y-Y_2)>0]-P[(X-X_1)(Y-Y_2)<0]), \tag{2.20}$$

where $(X,Y)^T, (X_1,Y_1)^T$ and $(X_2,Y_2)^T$ are independent copies.

Defining $R_i(x)$ as the rank of x_i in the sample $x=(x_1,\ldots,x_n)$ sorted (in ascending order) and, by analogy, $R_i(y)$ as the rank of y_i in the sample $y=(y_1,\ldots,y_n)$ sorted (in ascending order), Spearman's rho (2.20) can be estimated by

$$\hat{\rho}_S(X,Y) = \frac{\sum_{i=1}^{n}(R_i(x)-\overline{R_i}(x))(R_i(y)-\overline{R_i}(y))}{\sqrt{\sum_{i=1}^{n}(R_i(x)-\overline{R_i}(x))^2\sum_{i=1}^{n}(R_i(y)-\overline{R_i}(y))^2}} = 1-6\frac{\sum_{i=1}^{n}(R_i(x)-R_i(y))^2}{n(n^2-1)}. \tag{2.21}$$

As in the case of Kendall's tau, Spearman's rho can also be expressed as a function of a copula.

Theorem 2.3.3. *Let $(X,Y)^T$ be a vector of continuous random variables with copula C and product copula Π. Then, Spearman's rho for $(X,Y)^T$ is given by*

$$\rho_S(X,Y) = 3Q(C,\Pi) = 12\iint_{I^2} uv\,\mathrm{d}C(u,v)-3 = 12\iint_{I^2} C(u,v)\,\mathrm{d}u\mathrm{d}v-3. \tag{2.22}$$

Expression (2.22) is the expected value of the product of $U=F(x), V=G(y)\sim U(0,1)$ ($X\sim F$ and $Y\sim G$) whose joint distribution is given by the copula C. Therefore, (2.22) can be rewritten as

$$\rho_S(X,Y) = 12\iint_{I^2} C(u,v)\,\mathrm{d}u\mathrm{d}v-3 = 12\,\mathrm{E}[UV]-3 = \rho(U,V) = \rho(F(X),G(Y)), \tag{2.23}$$

where ρ is Pearson's correlation coefficient given in (2.12).

Similarly to the linear correlation coefficient:

- If $\tau(X,Y)>0$ or $\rho_S(X,Y)>0$, then X and Y are positively correlated.
- If $\tau(X,Y)<0$ or $\rho_S(X,Y)<0$, then X and Y are negatively correlated.
- If $\tau(X,Y)=0$ or $\rho_S(X,Y)=0$, then X and Y are linearly independent (but not necessarily independent).

2.3.3 Tail Dependence

The dependence between the variables in the upper-right quadrant and in the lower-left quadrant of I^2 is extremely relevant when we want to study the dependence between extreme values. This dependence is known as *tail dependence* and it will be a copula property. Therefore, it is invariant under strictly increasing transformations of the two random variables; see [Nelsen, 2006] or [Embrechts *et al*, 2003].

Definition 2.3.5. Let $(X,Y)^T$ be a vector of continuous random variables with marginal distributions functions F and G. The coefficient of upper tail dependence of $(X,Y)^T$ is

$$\lambda_U = \lim_{u\to 1^-} P[Y > G^{-1}(u) \mid X > F^{-1}(u)] \tag{2.24}$$

provided that the limit $\lambda_U \in I$ exists. Analogously, the coefficient of lower tail dependence of $(X,Y)^T$ is

$$\lambda_L = \lim_{u\to 0^+} P[Y \leq G^{-1}(u) \mid X \leq F^{-1}(u)] \tag{2.25}$$

provided that the limit $\lambda_L \in I$ exists.

The coefficients (2.24) and (2.25) can also be defined in terms of a copula function.

Definition 2.3.6. If a bivariate copula C with F and G defined as before is such that

$$\lambda_U = \lim_{u\to 1^-} \frac{1-2u+C(u,u)}{1-u} = \lim_{u\to 1^-} \frac{\overline{C}(u,u)}{1-u} \tag{2.26}$$

exists, then C has **upper tail dependence** if $\lambda_U \in (0,1]$. Similarly, if

$$\lambda_L = \lim_{u\to 0^+} \frac{C(u,u)}{u} \tag{2.27}$$

exists, C has **lower tail dependence** if $\lambda_L \in (0,1]$.

The tail dependence coefficients satisfy the following properties:

1. $\lambda_U, \lambda_L \in I$.
2. If $\lambda_U = 0$, then X and Y are said to be *asymptotically independent* in the upper tail. Otherwise, they are *asymptotically dependent* in the upper tail.
3. If $\lambda_L = 0$, then X and Y are said to be *asymptotically independent* in the lower tail. Otherwise, they are *asymptotically dependent* in the lower tail.

One needs to take into account that (2.26) and (2.27) can only be used when the copula have closed form expressions, such as in the case of Archimedean copulas and not in the case of Elliptical copulas, as we will later see. Otherwise, especially in the latter case, the tail dependence coefficients have to be determined using (2.24) and (2.25) and the following conditional probabilities

$$\lambda_U = \lim_{u\to 1^-} (P[V > u \mid U = u] + P[U > u \mid V = u]), \tag{2.28}$$

$$\lambda_L = \lim_{u\to 0^+} (P[V < u \mid U = u] + P[U < u \mid V = u]), \tag{2.29}$$

where $U,V \sim U(0,1)$ is a pair of random variables whose joint distribution is given by the copula C, $P[V \leq v \mid U = u] = \frac{\partial C(u,v)}{\partial u}$ and $P[V > v \mid U = u] = 1 - \frac{\partial C(u,v)}{\partial u}$ and, analogously, when conditioning on V; see [Embrechts *et al*, 2003].

There are a variety of copula classes, some of which can be found in [Joe, 2014]. However, in this thesis we will only consider two of the most used and important ones, the Elliptical and the Archimedean classes.

2.4 Elliptical Copulas

The class of elliptical copulas not only shares a set of properties with the multivariate normal distributions, but also enables the modelling of multivariate extremes and dependencies that differ from the ones of the Normal distribution. The elliptical copulas result from Sklar's theorem and from the fact that C has an elliptical distribution. However, the marginals may or may not follow an elliptical distribution; see [Embrechts *et al*, 2003], [Shemyakin and Kniazev, 2017]. Yet, the class of elliptical copulas does not have closed form expressions and is restricted to the fact that the copulas have radial symmetry, that is $C(u,v) = \overline{C}(u,v)$.

This class includes two of the most used copulas in applications, the Gaussian (or Normal) and the Student t.

2.4.1 Elliptical Distributions

[Embrechts *et al*, 2003] summarises in detail all the characteristics of an elliptical distribution. We will only introduce the more relevant for the aim of this thesis.

Definition 2.4.1. If $\mathbf{X}$ is a n-dimensional random vector and, for some $\mu \in \mathbb{R}^n$ and some $n \times n$ nonnegative definite symmetric matrix Σ, the characteristic function $\varphi_{\mathbf{X}-\mu}(\mathbf{t})$ of $\mathbf{X} - \mu$ is a function of the quadratic form $\mathbf{t}^T \Sigma \mathbf{t}$, $\varphi_{\mathbf{X}-\mu}(\mathbf{t}) = \phi(\mathbf{t}^T \Sigma \mathbf{t})$, then we say $\mathbf{X}$ has an **elliptical distribution** with parameters μ, Σ and ϕ, and we write $\mathbf{X} \sim E_n(\mu, \Sigma, \phi)$.

ϕ is known as the characteristic generator of an elliptical distribution.

If $\mathbf{X}$ is a 2-dimensional random vector with an elliptical distribution, $\mathbf{X} \sim E_2(\mu, \Sigma, \phi)$, where the matrix Σ is diagonal, if $\mathrm{Var}(X_i)$ is finite, then $\mathbf{X}$ has uncorrelated components. Also, if $\mathbf{X}$ has independent components, then $\mathbf{X} \sim N_2(\mu, \Sigma)$. Note that, if $\mathbf{X}$ follows a bivariate normal distribution, its characteristic function is given by

$$\Phi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{ i\mathbf{t}^t \mu - \frac{1}{2}\mathbf{t}^t \Sigma \mathbf{t} \right\} \tag{2.30}$$

and so, according to Definition 2.4.1, with $\phi(u) = \exp\left\{\frac{u}{2}\right\}$ and $u = \mathbf{t}^t \Sigma \mathbf{t}$, $\mathbf{X} \sim E_2(\mu, \Sigma, \phi)$.

[Embrechts *et al*, 2003] states that, whenever $0 < \mathrm{Var}(X_i), \mathrm{Var}(X_j) < \infty$, the matrix Σ is the variance-covariance matrix. Thus,

$$\rho(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i), \mathrm{Var}(X_j)}} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}, \tag{2.31}$$

where ρ is the linear correlation coefficient. This is the reason why ρ is a natural measure of dependence between random variables with a joint elliptical distribution.

One drawback of this class of joint distributions is that the margins have to be elliptical as well. To overcome this problem, one may choose to construct an elliptical copula with arbitrary margins (elliptical or not). However, in this case, Σ is no longer estimated by the variance-covariance matrix. [Embrechts *et al*, 2003] proposes an estimator for the linear correlation matrix of $\mathbf{X}$, R, whose components are given by $R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$. This estimator,

$$\hat{R}_{ij} = \sin\left(\frac{\pi \hat{\tau}(X_i, X_j)}{2}\right), \tag{2.32}$$

is robust where we must consider the non-parametric estimator $\hat{\tau}(X_i, X_j)$ of $\tau(X_i, X_j)$. (2.32) is an efficient estimator of R for elliptical copulas with or without elliptical marginals.

The density functions of the class of bivariate elliptical distributions $Q_\rho(s,t)$ are given by

$$q_\rho(s,t) = \frac{k^2}{\sqrt{1-\rho^2}} g\left(\frac{s^2 - 2\rho st + t^2}{1-\rho^2}\right), \tag{2.33}$$

where $\rho \in (-1, 1)$ is the linear correlation coefficient, $g : \mathbb{R} \to \mathbb{R}^+$ is such that $\int_{-\infty}^{+\infty} g(t)\,\mathrm{d}t < \infty$, and k is the normalising constant; see [Shemyakin and Kniazev, 2017].

This class includes distributions such as the Normal and the Student t.

2.4.2 The Gaussian Copula

The Gaussian, or Normal, Copula is obtained if (2.4) is constructed with a distribution function of a standardised bivariate normal distribution, which is denoted by $\Phi_\rho(s,t)$, where ρ is the correlation between the components. The *Gaussian Copula* is then defined by

$$C_\rho(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)), \tag{2.34}$$

where $\Phi^{-1}(t)$ is the inverse of the standard normal cumulative distribution function.

If we set $s = \Phi^{-1}(u)$ and $t = \Phi^{-1}(v)$, the Gaussian Copula density can be written as

$$c_\rho(u,v) = \frac{\phi_\rho(s,t)}{\phi(s)\phi(t)}, \tag{2.35}$$

where $\phi_\rho(s,t)$ is the density function of the standardised bivariate normal distribution with correlation ρ and $\phi(t)$ is the density of the standard normal distribution.

Since the expressions of $\phi_\rho(s,t)$ and $\phi(t)$ are known, one can rewrite (2.35) explicitly as

$$c_\rho(u,v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{\rho^2 s^2 + \rho^2 t^2 - 2\rho st}{2(1-\rho^2)}\right\}. \tag{2.36}$$

As said before, one can have arbitrary marginal distributions combined with the Gaussian Copula. Its goal is to transform the random variables X and Y, with distribution functions F and G, respectively, into standard normal variables through $S = \Phi^{-1}(F(X))$ and $T = \Phi^{-1}(G(Y))$. The dependence of X and Y will be then expressed in terms of the dependence structure of the standard normal variables. Therefore, we are expressing a nonlinear dependence between X and Y through the linear correlation coefficient of their standard normal transforms, S and T; see [Shemyakin and Kniazev, 2017].

Tail Dependence

This copula, since it belongs to the elliptical family, has radial symmetry, that is $C(u,v) = \overline{C}(u,v)$, and so the coefficient of the upper and lower tails will be the same. The Gaussian Copula is known to have no tail dependence as we can see in Fig. 2.1. Thus, $\lambda_U = \lambda_L = 0$.

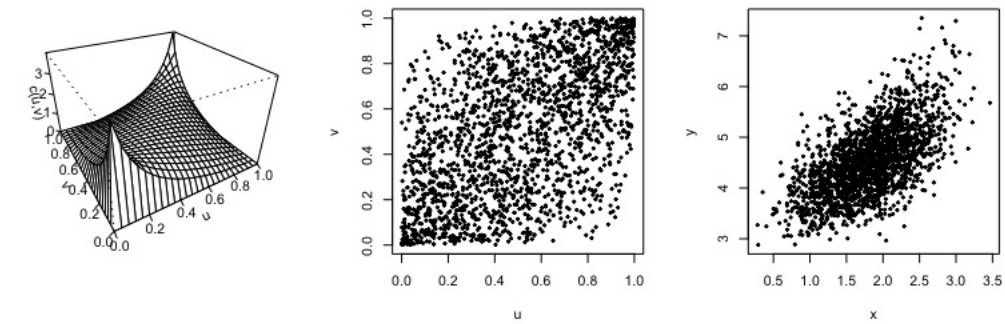


Fig. 2.1: Density of a Gaussian Copula with $\rho = 0.6$ (left); scatterplot of $(u,v)^T$, where $u = F(x)$ and $v = G(y)$ (middle) and scatterplot of $(x,y)^T$ where $X \sim Weibull(4,2)$ and $Y \sim LN(2.5, 0.15^2)$ (right). The plots were done by the author using the package *copula* in R; see [Yan, 2007].

Dependence Measures

For this copula, and as stated in (2.32), Kendall's concordance τ can be expressed as

$$\tau = \frac{2}{\pi}\arcsin(\rho). \tag{2.37}$$

Spearman's rho can be expressed as

$$\rho_S = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right) \tag{2.38}$$

and the linear correlation coefficient, as a function of ρ_S, is expressed as

$$\rho = 2\sin\left(\frac{\pi\rho_S}{6}\right), \tag{2.39}$$

and similarly for τ.

Simulation

If we want to estimate $H(x_0,y_0) = P[X \leq x_0, Y \leq y_0]$, where $X \sim F(x)$, $Y \sim G(y)$ and $H(x,y)$ as defined in (2.3), we can perform the following procedure described in [Shemyakin and Kniazev, 2017] as follows:

1. Generate independently two standard normal variables $z_1, z_2 \sim N(0,1)$.
2. Define correlated standard normal variables as $w_1 = z_1$ and $w_2 = \rho z_1 + \sqrt{1-\rho^2} z_2$.
3. Set $u = \Phi(w_1)$ and $v = \Phi(w_2)$.
4. Set $x = F^{-1}(u)$ and $y = G^{-1}(v)$. Note that implementations of this step depend on the distributions $F(x)$ and $G(y)$.

In order to obtain a sample (x_i, y_i), $i = 1,\dots,n$ from $C_\rho(F(x), G(y))$, we have to repeat the procedure above n times. $H(x_0,y_0)$ is estimated by the proportion of the sample elements which satisfy the condition $x_i \leq x_0$, $y_i \leq y_0$.

2.4.3 The Student *t*-Copula

When the interest focuses on modelling data which exhibits heavy-tailed behaviour, the Student *t*-Copula may be used instead of the Gaussian. [Demarta and McNeil, 2005] take special attention to its extremal properties, present some more flexible extensions of the copula and describe copulas related to the *t*-Copula through extreme value theory.

Multivariate *t* Distribution

The *univariate density* with $\eta > 0$ degrees of freedom, t_η, is defined by

$$t_\eta(x) = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\Gamma\left(\frac{\eta}{2}\right)\sqrt{\pi\eta}} \left(1 + \frac{x^2}{\eta}\right)^{\frac{-(\eta+1)}{2}} {}^{(b)}, \quad x \in \mathbb{R}. \tag{2.40}$$

The *d−variate density*, $t_{d,\eta}$, with correlation matrix R and η degrees of freedom of a random vector $\mathbf{x} \in \mathbb{R}^d$, is defined by

$$t_{d,\eta R}(\mathbf{x}) = |R|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\eta+d}{2}\right)}{\Gamma\left(\frac{\eta}{2}\right)(\pi\eta)^{\frac{d}{2}}} \left(1 + \frac{\mathbf{x}^T R^{-1} \mathbf{x}}{\eta}\right)^{\frac{-(\eta+d)}{2}}. \tag{2.41}$$

Therefore, the bivariate case with η degrees of freedom, where

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{2.42}$$

and $-1 < \rho < 1$, is given by

$$t_{2,\eta\rho}(x,y) = (1-\rho^2)^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\eta+2}{2}\right)}{\Gamma\left(\frac{\eta}{2}\right)(\pi\eta)} \left(1 + \frac{x^2 + y^2 - 2\rho xy}{\eta(1-\rho^2)}\right)^{\frac{-(\rho+2)}{2}}. \tag{2.43}$$

(b) $\Gamma(x) = \int_0^\infty z^{x-1} e^{-z} \mathrm{d}z, x > 0$ is the Gamma function

For simplicity, we will denote the bivariate case just as $t_{\eta\rho}(x,y)$ instead of $t_{2,\eta\rho}(x,y)$.

The following expression for the Student t-Copula is obtained constructing (2.4) with a bivariate t distribution function, $T_{\eta\rho}$

$$C_{\eta\rho}(u,v) = T_{\eta\rho}(T_\eta^{-1}(u), T_\eta^{-1}(v)), \tag{2.44}$$

where $T_\eta^{-1}(t)$ is the inverse of the univariate t distribution function with η degrees of freedom.

Once again, if we set $s = T_\eta^{-1}(u)$ and $r = T_\eta^{-1}(v)$, the density of the Student t-Copula can be written as

$$c_\rho(u,v) = \frac{t_{\eta\rho}(s,r)}{t_\eta(s)t_\eta(r)}, \tag{2.45}$$

where t_η and $t_{\eta\rho}$ are given in (2.40) and (2.41), respectively. Then, (2.45) can explicitly be written as

$$c_{\eta\rho}(u,v) = \frac{\Gamma\left(\frac{\eta+2}{2}\right)\Gamma\left(\frac{\eta}{2}\right)}{\sqrt{1-\rho^2}\ \Gamma^2\left(\frac{\eta+1}{2}\right)} \frac{\left(\left(1+\frac{s^2}{\eta}\right)\left(1+\frac{r^2}{\eta}\right)\right)^{\frac{\eta+1}{2}}}{\left(1+\frac{s^2+r^2-2\rho sr}{\eta(1-\rho^2)}\right)^{\frac{\eta+2}{2}}}. \tag{2.46}$$

As in the case of the Gaussian Copula, one can combine the Student t-Copula with any marginal distributions.

Tail Dependence

As the univariate and multivariate Student-t distributions, the Student t-Copula is characterised for having heavy tails. Therefore, this copula has tail dependence, as we can observe in Fig. 2.2. Additionally, due to its radial symmetry, the lower and upper coefficients are equal and given by

$$\lambda_L = \lambda_U = 2T_{\eta+1}(t), \tag{2.47}$$

where $T_{\eta+1}$ is the univariate t distribution function with $\eta+1$ degrees of freedom and

$$t = -\sqrt{\eta+1}\sqrt{\frac{1-\rho}{1+\rho}}. \tag{2.48}$$

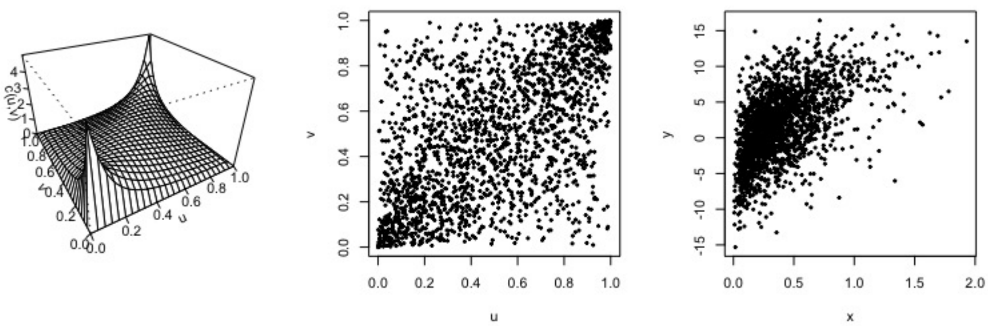


Fig. 2.2: Density of a Student t-Copula with $\rho = 0.6$ and $\eta = 5$ degrees of freedom (left); scatterplot of $(u,v)^T$, where $u = F(x)$ and $v = G(y)$ (middle) and scatterplot of $(x,y)^T$ where $X \sim Gamma(2,5)$ and $Y \sim N(2,5^2)$ (right). The plots were done by the author using the package *copula* in R; see [Yan, 2007].

Concordance Measures

For this copula, as said in (2.32), Kendall's concordance τ can be expressed as

$$\tau = \frac{2}{\pi}\arcsin(\rho). \tag{2.49}$$

Note that, for the Gaussian and Student t Copulas, τ is expressed in the same way and in terms of the correlation parameter ρ.

Simulation

We can as well estimate $H(x_0,y_0)$, but with a slightly modification of the procedure used for the Gaussian Copula. We recall that, if $Z \sim N(0,1)$ is a standard normal variable and $S \sim \chi^2(\nu)$ a chi-squared variable with ν degrees of freedom independent from Z, then

$$T = Z\sqrt{\frac{\nu}{S}} \tag{2.50}$$

follows a t distribution with ν degrees of freedom.

Hereupon, we can perform the following procedure, as described in [Shemyakin and Kniazev, 2017]:

1. Generate independently two standard normal variables $z_1, z_2 \sim N(0,1)$.
2. Generate a random variable s from $\chi^2(\nu)$ independent from z_1, z_2.
3. Define correlated standard normal variables $w_1 = z_1$ and $w_2 = \rho z_1 + z_2\sqrt{1-\rho^2}$.
4. Set $t_1 = w_1\sqrt{\frac{\nu}{s}}$ and $t_2 = w_2\sqrt{\frac{\nu}{s}}$.
5. Set $u_1 = T_\nu(t_1)$ and $u_2 = T_\nu(t_2)$.
6. Set $x = F^{-1}(u_1)$ and $y = G^{-1}(u_2)$. Note that implementations of this step depend on the distributions $F(x)$ and $G(y)$.

Again, we have to repeat the procedure n times in order to obtain a sample (x_i,y_i), $i = 1,\dots,n$ from $C_{\nu\rho}(F(x),G(y))$ and the sample elements which satisfy $x_i \le x_0$, $y_i \le y_0$ will be used to estimate $H(x_0,y_0)$.

For illustration of this procedure, let us consider the following example:

Example 1. Let $X \sim$ Gamma(2,5), $Y \sim N(2,5^2)$ and their dependence be modelled by a Student t-Copula, $C(u,v)$, with correlation parameter, $\rho = 0.6$, and $\eta = 5$ degrees of freedom. Our goal is to simulate the copula and estimate $H(3,1)$.

Considering a sample of size 2000, the number of elements satisfying $x_i \le 3$, $y_i \le 1$ is 858. It is straightforward to check that the required probability is 0.429.

If we simulate a Student t-Copula directly using the package *copula* in R, with the same marginal distributions and sample size, we obtain $H(3,1) = 0.4207$, which is very close. In Fig. 2.3 we compare the copula simulated by the procedure above (in green) and the copula simulated directly from the package (in blue). We can see, from the two scatterplots on the right, that both copulas have approximately the same features. It is also clear the dependence on the tails.

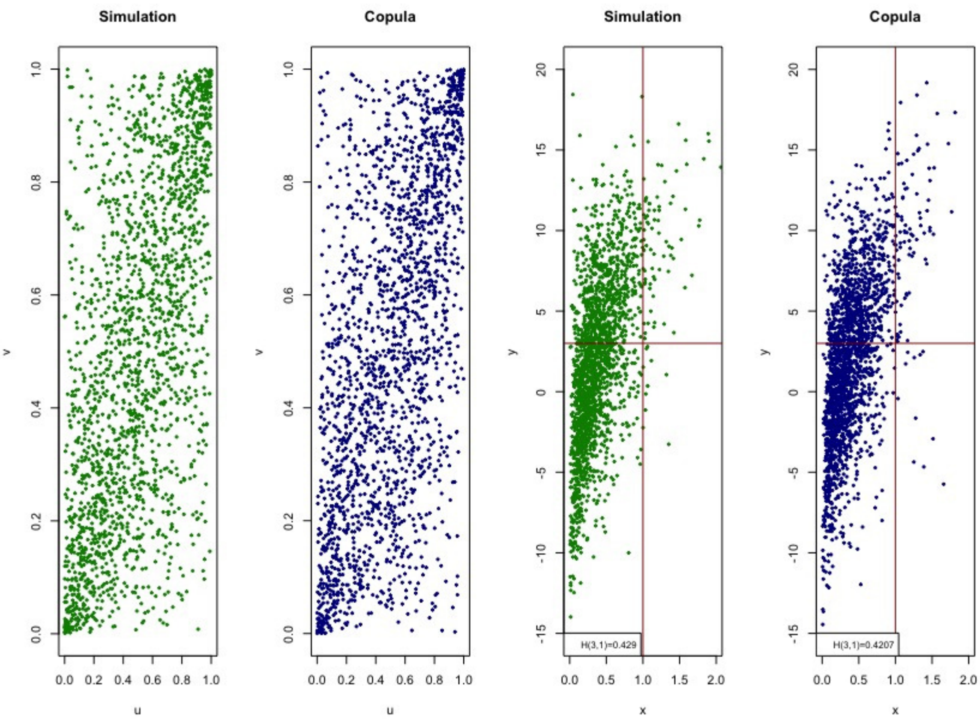


Fig. 2.3: Simulated Student t-Copula following the procedure proposed (green) vs simulated Student t-Copula using the package *copula* in R (blue).

2.5 Archimedean Copulas

The class of Archimedean copulas is one of the most important and used class of copulas for a number of reasons. Contrarily to the class of elliptical copulas, Archimedean copulas have closed form expressions. Also, many parametric families of copulas are Archimedean and this class allows a great variety of dependence structures. However, these copulas do not derive from Sklar's Theorem, which requires the need to introduce some conditions in order to considered them as copulas; see [Embrechts *et al*, 2003],

2.5.1 Definitions

Let X and Y be continuous random variables. The product copula $\Pi(u,v) = uv$, where $u = F(x)$ and $v = G(y)$ are the marginals of the copula, is the only instance in which the joint distribution function factors into a product of a function of the marginals. But with a certain deviation from this copula, one may have another way of constructing copulas. Whenever it is possible to write $\lambda(H(x,y)) = \lambda(F(x))\lambda(G(y))$ for a function λ, which has to be positive on $(0,1)$, through the transformation $\varphi(t) = -\log(\lambda(t))$, we can also write $H(x,y)$ as a sum of functions of the marginals, u and v. Thus,

$$\varphi(H(x,y)) = \varphi(F(x)) + \varphi(G(y)) \tag{2.51}$$

or, in terms of copulas,

$$\varphi(C(u,v)) = \varphi(u) + \varphi(v). \tag{2.52}$$

Moreover, since the interest relies on having a form of constructing copulas, we need to introduce the concept of pseudo-inverse $\varphi^{[-1]}$ so that we can have $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$; see [Nelsen, 2006], [Shemyakin and Kniazev, 2017].

Definition 2.5.1. Let φ be a continuous, strictly decreasing function from I to $[0,\infty]$ such that $\varphi(1)=0$. The **pseudo-inverse** of φ is the function $\varphi^{[-1]}:[0,\infty]\to I$ given by

$$\varphi^{[-1]}(t)=\begin{cases}\varphi^{-1}(t), & 0\le t\le\varphi(0)\\ 0, & \varphi(0)\le t\le\infty\end{cases} \tag{2.53}$$

Note that $\varphi^{[-1]}$ is continuous, nonincreasing on $[0,\infty]$, and strictly decreasing on $[0,\varphi(0)]$. Furthermore, $\varphi^{[-1]}(\varphi(u))=u$ in I and

$$\varphi(\varphi^{[-1]}(t))=\begin{cases}t, & 0\le t\le\varphi(0)\\ \varphi(0), & \varphi(0)\le t\le\infty\end{cases} \tag{2.54}$$

Finally, if $\varphi(0)=\infty$, then $\varphi^{[-1]}=\varphi^{-1}$.

Pseudo-inverses are useful to extend the inverse functions to functions of limit range; see [Embrechts *et al*, 2003].

Theorem 2.5.1. *Let φ be a continuous, strictly decreasing function from I to $[0,\infty]$ such that $\varphi(1)=0$, and let $\varphi^{[-1]}$ be the pseudo-inverse of φ. Let C be the function from I^2 to I given by*

$$C(u,v)=\varphi^{[-1]}(\varphi(u)+\varphi(v)). \tag{2.55}$$

Then C is a copula if and only if φ is convex.

For the proof see [Nelsen, 2006].

Definition 2.5.2. A function $g:\mathbb{R}^n\to\mathbb{R}$ is convex if its domain is a convex set and for all x, y in its domain, and all $t\in[0,1]$,we have

$$g(tx+(1-t)y)\le tg(x)+(1-t)g(y). \tag{2.56}$$

See [Boyd and Vandenberghe, 2004].

If a copula $C(u,v)$ is of the form (2.55) and the function φ is convex, then $C_\alpha(u,v)$ is called an *Archimedean copula* and φ its generator.

If the second derivative $\varphi''(t)$ exists, then the *density of Archimedean copulas* can be expressed as

$$c(u,v)=-\frac{\varphi''(C(u,v))\varphi'(u)\varphi'(v)}{(\varphi'(C(u,v)))^3}. \tag{2.57}$$

On this thesis we will be focusing on the three most frequent copulas of this class, which are the Clayton, Frank and Gumbel-Hougaard Copulas.

Tab. 2.1: Three Archimedean Copulas with association parameter α and some of theirs properties.

Copula	$\varphi_\alpha(t)$	$\varphi_\alpha^{[-1]}(s)$	$\alpha\in$	Limiting and Special Cases
Clayton	$\frac{1}{\alpha}(t^{-\alpha}-1)$	$\max\{(1+\alpha s)^{-1/\alpha},0\}$	$[-1,\infty)\setminus\{0\}$	$C_{-1}=W, C_0=\Pi, C_\infty=M$
Frank	$-\log(\frac{e^{-\alpha t}-1}{e^{-\alpha}-1})$	$-\frac{1}{\alpha}\log([1+e^{-s}(e^{-\alpha}-1)])$	$\mathbb{R}\setminus\{0\}$	$C_{-\infty}=W, C_0=\Pi, C_\infty=M$
Gumbel-Hougaard	$(-\log(t))^\alpha$	$e^{-s^{1/\alpha}}$	$[1,\infty)$	$C_1=\Pi, C_\infty=M$

2.5.2 Properties

Theorem 2.5.2. *Let C be an Archimedean copula with generator φ. Then*

1. *C is symmetric, i.e. $C(u,v) = C(v,u)$ for all $u,v \in I$.*

2. *C is associative, i.e. $C(C(u,v),w) = C(u,C(v,w))$ for all $u,v,w \in I$.*

3. *For every constant, $c \geq 0$, $c\varphi$ is also a generator of C.*

See [Embrechts *et al*, 2003] for the proof of 1 and 2. [Embrechts *et al*, 2003] also showed that the associativity property of this class of copulas is not shared by copulas in general.

2.5.3 Tail Dependence

When we are in the class of Archimedean copulas, tail dependence is easily expressed in terms of the generator of each copula.

Theorem 2.5.3. *Let C be an Archimedean copula with generator φ. If $\frac{\partial \varphi^{-1}}{\partial t}(0)$ is finite, then*

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \tag{2.58}$$

does not have upper tail dependence. If C has upper tail dependence, then $\frac{\partial \varphi^{-1}}{\partial t}(0) = -\infty$ and its coefficient is given by

$$\lambda_U = 2 - 2 \lim_{s \to 0} \frac{\frac{\partial \varphi^{-1}}{\partial t}(2s)}{\frac{\partial \varphi^{-1}}{\partial t}(s)}. \tag{2.59}$$

Theorem 2.5.4. *Let C be an Archimedean copula with generator φ. If φ is as in Theorem 2.5.3, then the coefficient of lower tail dependence of the copula $C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v))$ is equal to*

$$\lambda_L = 2 \lim_{s \to \infty} \frac{\frac{\partial \varphi^{-1}}{\partial t}(2s)}{\frac{\partial \varphi^{-1}}{\partial t}(s)}. \tag{2.60}$$

For the proof see [Embrechts *et al*, 2003].

The variety of dependence structures of the copulas of the Archimedean class can be seen on Tab. 2.2. Likewise the Gaussian Copula, the Frank one does not show any tail dependence, while the Clayton Copula shows lower tail dependence and the Gumbel-Hougaard upper tail dependence.

Tab. 2.2: Tail dependence for three Archimedean copulas.

Copula	$\alpha \in$	λ_L	λ_U
Clayton	$[-1,\infty) \setminus \{0\}$	$2^{-\frac{1}{\alpha}}$	0
Frank	$\mathbb{R} \setminus \{0\}$	0	0
Gumbel-Hougaard	$[1,\infty)$	0	$2 - 2^{\frac{1}{\alpha}}$

2.5.4 Concordance Measures

In the case of an Archimedean copula, Kendall's tau can be determined in a easier way since it can be expressed as an integral of its generator and its derivative. If X and Y are random variables and its dependence is modelled by an Archimedean copula C generated by φ, Kendall's tau of X and Y is given by

$$\tau_C = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} \, \mathrm{d}t. \tag{2.61}$$

For the proof see [Embrechts *et al*, 2003].

Equation (2.61) can also be used as a way of estimating the association parameter α, as we can observe on Tab. 2.3 below

Tab. 2.3: Kendall's tau and copula association parameter.

Copula	$\alpha \in$	τ	α
Clayton	$[-1,\infty)\setminus\{0\}$	$\frac{\alpha}{\alpha+2}$	$\frac{2\tau}{1-\tau}$
Frank	$\mathbb{R}\setminus\{0\}$	$1+\frac{4(D(\alpha)-1)}{\alpha}$	–
Gumbel-Hougaard	$[1,\infty)$	$\frac{\alpha-1}{\alpha}$	$\frac{1}{1-\tau}$

where $D(\alpha) = \frac{1}{\alpha}\int_0^\alpha \frac{t}{e^t-1}\,\mathrm{d}t$ is the Debye's integral.

The Clayton Copula

With the generator and pseudo-inverse defined on Tab. 2.1, the *Clayton Copula* is given by

$$C_\alpha(u,v) = \max\{(u^{-\alpha}+v^{-\alpha}-1)^{-\frac{1}{\alpha}},0\}, \tag{2.62}$$

with $\alpha \in [-1,\infty)\setminus\{0\}$. However, this copula is typically applied when $\alpha > 0$ and so (2.62) can be written as

$$C_\alpha(u,v) = (u^{-\alpha}+v^{-\alpha}-1)^{-\frac{1}{\alpha}}. \tag{2.63}$$

Thus, its density is given by

$$c_\alpha(u,v) = \frac{(\alpha+1)(uv)^\alpha}{(u^\alpha+v^\alpha-(uv)^\alpha)^{\frac{1}{\alpha}+2}}. \tag{2.64}$$

As said before, this copula has lower tail dependence, so it is used when the dependence between low values of the marginals is stronger than the dependence between values close to 1.

Although not as used in literature as the Survival Gumbel Copula, the *Survival Clayton Copula* is given by

$$\overline{C}_\alpha(u,v) = u+v-1+((1-u)^{-\alpha}+(1-v)^{-\alpha}-1)^{-\frac{1}{\alpha}}, \tag{2.65}$$

with $\alpha > 0$.

The Frank Copula

With the generator and pseudo-inverse defined on Tab. 2.1, the *Frank Copula* is given by

$$C_\alpha(u,v) = -\frac{1}{\alpha}\log\left(\frac{1-e^{-\alpha}-(1-e^{-\alpha u})(1-e^{-\alpha v})}{1-e^{-\alpha}}\right). \tag{2.66}$$

and its density is given by

$$c_\alpha(u,v) = \frac{\alpha(1-e^{-\alpha})e^{-\alpha(u+v)}}{(1-e^{-\alpha}-(1-e^{-\alpha u})(1-e^{-\alpha v}))^2}, \tag{2.67}$$

where $\alpha \neq 0$.

The Frank Copula does not have any tail dependence, so it is used when the strength of the dependence is relatively similar for all values of the marginals.

The Gumbel-Hougaard Copula

With the generator and pseudo-inverse defined on Tab. 2.1, the *Gumbel-Hougaard Copula* and its density are, respectively, given by,

$$C_\alpha(u,v) = \exp\{-((-\log(u))^\alpha + (-\log(v))^\alpha)^{\frac{1}{\alpha}}\}, \tag{2.68}$$

$$c_\alpha(u,v) = (uv)^{-1}(\log(u)\log(v))^{\alpha-1}(w^{\frac{2}{\alpha}-2} + (\alpha-1)w^{\frac{1}{\alpha}-2})C_\alpha(u,v), \tag{2.69}$$

where $w = (-\log(u))^\alpha + (-\log(v))^\alpha$ and $\alpha \geq 1$. For simplicity, we will refer to this copula as the Gumbel Copula.

In contrast to the Clayton Copula, the Gumbel Copula has upper tail dependence, so it is used when the dependence between high values of the marginals is stronger than the dependence between values close to 0.

In cases where there is a combination of Exponential and Weibull margins, the *Survival Gumbel Copula* has a lot of applications; see [Shemyakin and Kniazev, 2017]. It is given by

$$\overline{C}_\alpha(u,v) = u + v - 1 + \exp\{-((-\log(1-u))^\alpha + (-\log(1-v))^\alpha)^{\frac{1}{\alpha}}\}, \tag{2.70}$$

with $\alpha \geq 1$.

We can see in Fig. 2.4 the difference of the three copulas, mainly on the tail dependence. Clayton clearly shows a lower tail dependence, $\lambda_L = 0.7937$, while Gumbel has an upper coefficient of 0.6805. Frank does not have any as stated before, a fact which is reinforced by the scatterplot.

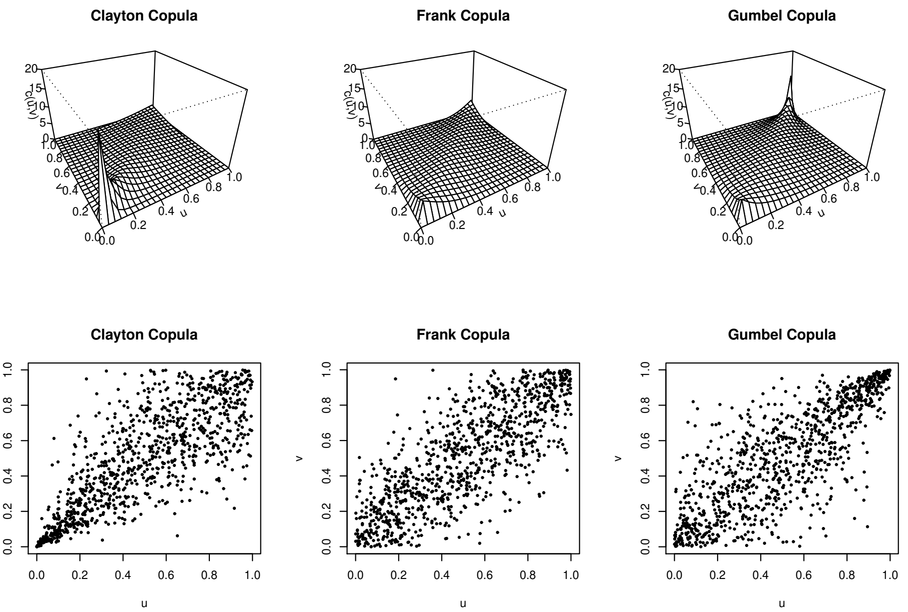


Fig. 2.4: Comparison between three Archimedean copulas with Kendall's tau τ=0.6 and 1000 observations generated. $\lambda_L^C = 0.7937$, $\lambda_U^G = 0.6805$ and $\lambda_U^C = \lambda_L^F = \lambda_U^F = \lambda_L^G = 0$. The plots were done by the author using the package *copula* in R; see [Yan, 2007].

2.5.5 Simulation

[Shemyakin and Kniazev, 2017] show that there exist convenient formulas for direct calculation of probabilities $H(x_0,y_0) = P[X \leq x_0, Y \leq y_0]$. However, it is required additional comprehension of the structure of the member of the Archimedean class.

For the simulation of Archimedean copulas, we can consider two different construction methods: one based on Kendall's distribution and the other based on Marshall-Olkin Construction.

Theorem 2.5.5. *Let C be an Archimedean copula generated by φ and let*

$$K_C(t) = V_C(\{(u,v) \in I^2 \mid C(u,v) \leq t\}).$$

Then, for any t in I,

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}. \tag{2.71}$$

See [Nelsen, 2006] for the proof. (2.71) is known as *Kendall's distribution function*.

Corollary 2.5.5.1. *If $(U,V)^T$ has distribution function C, where C is an Archimedean copula generated by φ, then the function K_C given by* (2.71) *is the distribution function of the random variable $C(U,V)$.*

Theorem 2.5.6. *Under the hypotheses of Corollary 2.5.5.1, the joint distribution function $H(s,t)$ of the random variables $S = \frac{\varphi(U)}{\varphi(U)+\varphi(V)}$ and $T = C(U,V)$ is given by $H(s,t) = sK_C(t)$ for all $(s,t) \in I^2$. Hence S and T are independent, and S is uniformly distributed on I.*

For the proof see [Embrechts *et al*, 2003].

Theorem 2.5.6 is on the base of one of the sampling procedures for Archimedean copulas. Assuming that F and G are the marginal distributions of the copula and the generator φ_α corresponds to an Archimedean copula with association parameter value α, the procedure is as stated in [Embrechts *et al*, 2003] and [Shemyakin and Kniazev, 2017]:

1. Generate independently two variables s and w, uniform in $[0,1]$.
2. Solve for $t = K_C^{-1}(w)$.
3. Set $u = \varphi_\alpha^{[-1]}(s\varphi_\alpha(t))$ and $v = \varphi_\alpha^{[-1]}((1-s)\varphi_\alpha(t))$.
4. Set $x = F^{-1}(u)$ and $y = G^{-1}(v)$.

Note that s and t correspond to S and T from Theorem 2.5.6.

Another sampling procedure for simulating Archimedean copulas was introduced by [Marshall and Olkin, 1988]. There exists a nonnegative random variable W such that the generator of many Archimedean copulas, $\varphi_\alpha(t)$, is the inverse of its moment generating function, $M(t) = \mathrm{E}[e^{-tW}]$, i.e. its Laplace transform. If S and T are two independent variables uniformly distributed in I,

$$U = M\left(-\frac{\log(S)}{W}\right), \qquad V = M\left(-\frac{\log(T)}{W}\right) \tag{2.72}$$

are also uniform on I, and their joint distribution is an Archimedean copula with generator $\varphi_\alpha(t)$, [Shemyakin and Kniazev, 2017]. If the distribution of W is easy to sample, one can use the following algorithm to sample Archimedean copulas.

1. Generate a copy of random variable w.
2. Draw s and t independently from $U[0,1]$.
3. Set $u = \varphi_\alpha^{[-1]}\left(-\frac{\log(s)}{w}\right)$ and $v = \varphi_\alpha^{[-1]}\left(-\frac{\log(t)}{w}\right)$.
4. Set $x = F^{-1}(u)$ and $y = G^{-1}(v)$.

For both procedures and, as in the simulation for elliptical copulas, to obtain a sample (x_i, y_i), $i = 1, \ldots, n$ from $C_\alpha(F(x), G(y))$ we have to repeat the algorithm n times.

3 | Statistics of Copulas

3.1 Estimation

The choice of the best distribution, either univariate, or multivariate, is generally not an easy task. As stated in [Huard *et al*, 2006], it is not trivial and it is linked to the estimation of the parameters. Additionally, if more parameters have to be estimated, more difficult this process becomes. The main question is how to estimate the copula parameters given a sample $(x_1, y_1), \ldots, (x_n, y_n)$ of (X, Y).

First, one has to decide if the parameters are estimated in a parametric, semiparametric or non-parametric way. Some authors, e.g. [Genest and Favre, 2007], choose to perform it non-parametrically, since the dependence captured by the copula is not related to the individual behaviour of the variables. Therefore, they only consider rank-based estimators. However, some authors; see for instance [Joe, 2014] and [Shemyakin and Kniazev, 2017], prefer to estimate the parameters parametrically.

3.1.1 Parametric Inference

[Joe, 2014] enumerates a set of advantages of using parametric inference methods for copula models. For instance, they are easier to implement numerically and can be used in high dimensions. From the point of view of [Shemyakin and Kniazev, 2017], parametric models generally are less data dependent and can have better predictive quality than other approaches.

The most general and used method of estimation for parametric models is based on the likelihood. Assume that we have an *iid* sample $\boldsymbol{z} = (x_1, y_1), \ldots, (x_n, y_n)$ of (X, Y). Let F and G be the distribution functions of X and Y and f and g the probability density functions of X and Y, respectively. The joint probability function of (X, Y) is given by

$$f(x, y \mid \boldsymbol{\theta}) = c(F(x \mid \boldsymbol{\alpha}_1), G(y \mid \boldsymbol{\alpha}_2) \mid \boldsymbol{\delta}) f(x \mid \boldsymbol{\alpha}_1) g(y \mid \boldsymbol{\alpha}_2), \tag{3.1}$$

and we can write the log-likelihood of the model as

$$L(\boldsymbol{\theta} \mid \boldsymbol{z}) = \sum_{i=1}^{n} \log(c(F(x_i \mid \boldsymbol{\alpha}_1), G(y_i \mid \boldsymbol{\alpha}_2) \mid \boldsymbol{\delta})) + \sum_{i=1}^{n} (\log(f(x_i \mid \boldsymbol{\alpha}_1)) + \log(g(y_i \mid \boldsymbol{\alpha}_2))), \tag{3.2}$$

where $\boldsymbol{\alpha}_1$ is the vector of parameters of the distribution of X, $\boldsymbol{\alpha}_2$ is the vector of parameters of the distribution of Y, $\boldsymbol{\delta}$ is the vector of the copula parameters and $\boldsymbol{\theta} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\delta})$ is the vector of all model parameters. It is straightforward that the first part of (3.2) is the log-likelihood of the bivariate copula and the second part is the log-likelihood of each marginal.

Maximum Likelihood Estimator

The maximum likelihood estimation (MLE) method can be used to estimate simultaneously the parameters of the marginal distributions and the copula parameters. The estimator is then obtained by maximising the function (3.2).

Assuming that the regularity conditions are satisfied, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is given by

$$\hat{\boldsymbol{\theta}}_{MLE} = \operatorname{argmax} L(\boldsymbol{\theta} \mid \boldsymbol{z}). \tag{3.3}$$

The estimator (3.3) is asymptotically normally distributed, *i.e.*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, I^{-1}(\boldsymbol{\theta}_0)), \tag{3.4}$$

where $\hat{\boldsymbol{\theta}}_{MLE} \xrightarrow{p} \boldsymbol{\theta}_0$, $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ and $I^{-1}(\boldsymbol{\theta}_0)$ is the Fisher information evaluated at $\boldsymbol{\theta}_0$; see [Joe, 2014].

Inference Functions for Margins

In the case where there exist too many parameters, (3.2) can be difficult to maximise. Therefore, in order to overcome this problem, [Joe, 1997] proposed a two-step method called *Inference Functions for Margins (IFM)*. The method consists, firstly on the estimation of the marginal parameters separately by maximum likelihood and, secondly on using them to estimate the copula parameters. To sum up:

1. We obtain by MLE the estimates of the parameters of the marginal distributions

$$\hat{\boldsymbol{\alpha}} = \text{argmax} \sum_{i=1}^{n} (\log(f(x_i \mid \boldsymbol{\alpha}_1)) + \log(g(y_i \mid \boldsymbol{\alpha}_2))), \tag{3.5}$$

where $\boldsymbol{\alpha}_1$ is the vector of the parameters of the distribution of X, $\boldsymbol{\alpha}_2$ is the vector of parameters of the distribution of Y and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$ is the vector of all the marginal parameters.

2. We obtain by MLE the estimates of the copula parameters

$$\hat{\boldsymbol{\delta}} = \text{argmax} \sum_{i=1}^{n} \log(c(F(x_i \mid \hat{\boldsymbol{\alpha}}_1), G(y_i \mid \hat{\boldsymbol{\alpha}}_2), \mid \boldsymbol{\delta})) \tag{3.6}$$

where $\boldsymbol{\delta}$ is the vector of copula parameters.

The IFM estimator is then given by the vector $\hat{\boldsymbol{\theta}}_{IFM} = (\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \hat{\boldsymbol{\delta}})$. Under certain regularity conditions, the IFM estimator is consistent and asymptotically normal. More about its asymptotic efficiency can be found in [Joe, 2005].

3.1.2 Semiparametric Inference

The main difference between the parametric estimation and the semiparametric one is that the latter does not make any assumption on the marginal distributions. Being so, we first estimate the marginal distributions by the empirical functions and then estimates the copula parameters through the maximum likelihood estimation.

Let $\hat{F}(x)$ be the empirical distribution of the random variable X, *i.e.*

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(X_i \leq x)}, \tag{3.7}$$

where $I_{(X_i \leq x)}$ is the indicator function which is equal to 1 when $X_i \leq x$, and 0 otherwise. After obtaining $\hat{u}_i = \hat{F}(x_i)$ and $\hat{v} = \hat{G}(y_i)$, we estimate the copula parameters by maximising

$$\sum_{i=1}^{n} \log(c(\hat{u}_i, \hat{v}_i) \mid \boldsymbol{\delta})) \tag{3.8}$$

and the maximum pseudo-likelihood estimator (MPLE) is given by

$$\hat{\boldsymbol{\theta}}_{MPLE} = \text{argmax} \sum_{i=1}^{n} \log(c(\hat{u}_i, \hat{v}_i) \mid \boldsymbol{\delta})). \tag{3.9}$$

Note that this method can only be implemented if the variables X and Y are continuous; see [Joe, 2014]. Moreover, [Genest *et al*, 1995a] proved that this estimator is consistent and asymptotically normal when the copula model is specified correctly. One of the advantages of this approach is that it can be useful for model comparison and selection, since it eliminates the possible misspecification of the marginal distributions; see [Shemyakin and Kniazev, 2017].

3.1.3 Non-parametric Inference

The last case of copula inference is the non-parametric approach. [Joe, 2014] shows that, even if the parameters are estimated in a parametric way, non-parametric methods can be useful to assess the adequacy of the copula model fit. We may have two methods of non-parametric copula inference, one based on the empirical copula, and the other based on moments.

Empirical Copula

Let us define the pseudo-observations $\boldsymbol{z}_1 = (z_{11}, z_{12}), \ldots, \boldsymbol{z}_n = (z_{n1}, z_{n2})$ of the pseudo-vector $\boldsymbol{Z}$ by

$$\boldsymbol{z}_i = \left(\frac{n\hat{F}(x_i)}{n+1}, \frac{n\hat{G}(y_i)}{n+1} \right), \quad i = 1, \ldots, n, \tag{3.10}$$

with $\hat{F}(x_i)$ and $\hat{G}(y_i)$ defined as before. The pseudo-observations can be interpreted as a sample from the copula C. The *empirical copula* C_n is then given by

$$C_n(u,v) = \frac{1}{n} \sum_{i=1}^{n} I_{[0,1]^2}(Z_{i1} \leq u, Z_{i2} \leq v), \tag{3.11}$$

where $I_{[0,1]^2}(Z_{i1} \leq u, Z_{i2} \leq v)$ is the indicator function, which take the value 1 if $(Z_{i1} \leq u, Z_{i2} \leq v) \in [0,1]^2$, and 0 otherwise. Therefore, one can estimate the copula C by (3.11). Moreover, it is a consistent estimator of C. We can also compute the mean relative error $\left(\left| \frac{C - C_n}{C} \right| \right)$ of considering the copula estimator C_n instead of the true copula C; see [Hofert *et al*, 2018].

Method of Moments

The other type of non-parametric inference for copulas is based on the generalised method of moments. The main idea is to find first functional relationships between the copula parameters and some distribution characteristics, for example moments, then estimate distribution moments with sample moments and finally estimate the copula parameters by plugging in sample moments, instead of the usual distribution moments, into the relationships found in the first step; see [Shemyakin and Kniazev, 2017]. It is reinforced in [Genest and Favre, 2007] that, since the dependence structure does not depend on the individual behaviour of the variables X and Y, the estimators for the copula parameters should be based on the ranks of the observations, *i.e.*, rank-based estimators, instead of being parametrically estimated. Therefore, they consider two different cases of estimation, one based on Kendall's tau and the other based on Spearman's rho.

As we stated before, we can find a functional relationship between the copula parameter and Kendall's tau, through (2.17), a sample value $\hat{\tau}$ of τ. For the Archimedean copulas, this approach is straightforward using (2.61) and Tab 2.3. In the case of the Elliptical copulas, the relationship is the same for the Gaussian Copula and the Student t-Copula and it is given by (2.37).

Regarding the estimation of the copula parameters based on Spearman's Rho, the approach is the same. We can find a functional relationship between the parameter and Spearman's rho, through (2.21), a sample value $\hat{\rho}_S$ of ρ_S.

Note that, for two-parameter copulas, as it is the case of the Student t-Copula, it is less straightforward to know what moment measures should be used to estimate both parameters. For the Student t-Copula we can use Kendall's tau or Spearman's rho to estimate the correlation parameter ρ but not its degrees of freedom η; see [Shemyakin and Kniazev, 2017].

3.1.4 Bayesian Inference

In the classical approach to statistics, we consider a sample $(x_1, \ldots, x_n)$ of a random variable X which follows an unknown distribution $F(x)$ with a fixed vector of parameters $\boldsymbol{\theta}$, also unknown. Then we estimate $\boldsymbol{\theta}$, via maximum likelihood for example, obtain its confidence interval, and perform hypothesis testing on the model parameters. On the other hand, in the case of Bayesian inference, $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ is a model for the data while $\pi(\boldsymbol{\theta})$ is the distribution which we believe that the parameter follows, which is called the prior distribution. We can then obtain the posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$, through the Bayes' Theorem:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_\Theta f(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}, \quad \boldsymbol{\theta} \in \Theta, \tag{3.12}$$

where $\Theta = \{\boldsymbol{\theta} : \pi(\boldsymbol{\theta}) > 0\}$ is the support of $\boldsymbol{\theta}$ and $\int_\Theta f(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the normalising constant. The prior distribution gives us the strength of the belief in $\boldsymbol{\theta}$ before observing $\boldsymbol{x}$, while the posterior distribution represents the strength of the belief in $\boldsymbol{\theta}$ taking into account the observed data $\boldsymbol{x}$; see [Atique and Atooh-Okine, 2018].

After constructing the posterior distribution, we can obtain credible regions for the parameter which can be interpret as true probabilistic measures, contributing to one of the main advantages of the Bayesian approach towards the classical one. However, the choice of the prior distribution is one of main problems in Bayesian statistics and can be viewed as a drawback. In the case where there is no previous knowledge about the parameter, one can have non-informative priors. Otherwise, informative priors can be elicited by subjective considerations, or obtained via empirical Bayes approach using relevant data.

As stated in [Joe, 2014], there is no advantage on using Bayesian inference if there is no historical information that can be included in the prior, since the log-likelihood would dominate the prior, especially with large samples. On the other hand, some cases are enumerated in [Smith, 2011] where a Bayesian approach might be preferable, for instance when estimating the copula model, the objective generally lies on making inferences on measures of dependence and/or quantiles, and the evaluation of the posterior distribution of these quantities is straightforward using Markov Chain Monte Carlo methods. When the likelihood of the marginal and/or of the copula are hard to maximise, the Bayesian approach can be seen as an alternative to the IFM method.

As happens with the classical inference for copulas, one can have a fully Bayesian or a two-step estimation approaches. However, according to [Dos Santos Silva and Lopes, 2008], the two approaches lead, on average, to the same results. In addition, they affirmed that it is more desirable to jointly estimate all the parameters, in a Bayesian framework, in order to obtain a complete characterisation of the posterior distribution and to construct the dependence structure of all the variables.

As before, the joint probability function of (X,Y) is given by

$$f(x,y \mid \boldsymbol{\theta}) = c(F(x \mid \boldsymbol{\alpha}_1), G(y \mid \boldsymbol{\alpha}_2) \mid \boldsymbol{\delta}) f(x \mid \boldsymbol{\alpha}_1) g(y \mid \boldsymbol{\alpha}_2), \tag{3.13}$$

where c is the copula density, $\boldsymbol{\alpha}_1$ is the vector of parameters of the distribution of X, $\boldsymbol{\alpha}_2$ is the vector of parameters of the distribution of Y, $\boldsymbol{\delta}$ is the vector of the copula parameters and $\boldsymbol{\theta} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\delta})$ is the vector of all model parameters.

Let $\boldsymbol{z} = (x_1,y_1),\ldots,(x_n,y_n)$ be an *iid* sample of (X,Y), then the likelihood function is given by

$$L(\boldsymbol{\theta} \mid \boldsymbol{z}) = \prod_{i=1}^{n} c(F(x_i \mid \boldsymbol{\alpha}_1), G(y_i \mid \boldsymbol{\alpha}_2) \mid \boldsymbol{\delta}) \prod_{i=1}^{n} f(x_i \mid \boldsymbol{\alpha}_1) \prod_{i=1}^{n} g(y_i \mid \boldsymbol{\alpha}_2). \tag{3.14}$$

According to (3.12), the posterior distribution for the model is then given by

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{z}) \propto L(\boldsymbol{\theta} \mid \boldsymbol{z}) \pi(\boldsymbol{\theta}). \tag{3.15}$$

To complete the specification of the model one just has to define independent prior distributions for each parameter of $\boldsymbol{\theta}$, informative or non-informative. Bayesian inference can be a better alternative towards the classical inference if it provides a narrower range for the copula parameters.

3.2 Model Selection

The choice of an appropriate model to fit a data set can be difficult. For instance, [Huard *et al*, 2006] acknowledged that, for choosing the best bivariate model, one has to find the optimal marginal distributions first, and then the optimal copula.

The search for the best univariate distributions can be performed in a graphical way, with the help of histograms, QQ-plots, PP-plots and by using kernel distributions estimation methods, for example, or with the assistance of goodness-of fit tests, such as the Chi-Square (χ^2), the Kolmogorov-Smirnov (KS), the Cramér-von-Mises (CvM) and the Anderson-Darling (AD). However, in the case of unspecified distributions, i.e. distribution with unknown parameters which have to be estimated first, one has to take some precautions when performing some of these tests. One can also take into account some quality information criteria, such as Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) and log-likelihood.

Regarding to the choice of the copula, one can use the information criteria above, along with the Deviance Information Criterion (DIC), graphical tools and also perform goodness-of-fit tests. [Huard *et al*, 2006] propose a Bayesian copula selection which is independent of the parameter choice and estimation. Briefly, they define an exhaustive and mutually exclusive hypothesis and then compute the probability of each hypothesis when the observed data is given. They define prior probabilities for each hypothesis and compute the posterior probabilities. The hypothesis with the highest posterior probability is the "right" copula. In [Genest and Favre, 2007], some graphical tools for assessing the fit of the copula, such as the use of a K-Plot, are included.

The problem of overfitting, which occurs when a model fits well to a particular set of data but may fail to predict future events or to fit additional data, is considered in [Shemyakin and Kniazev, 2017]. Usually, it is a model which has more parameters than can be justified by the data available. It is also important to have a model with as few parameters as possible due to the fact that, not only it is harder to estimate a model over parameterised, but also it may become close enough to the observed data arbitrarily; see [Erntell, 2013]. Goodness-of-fit tests, contrary to information criteria as the AIC and BIC, do not address this problem.

3.2.1 Marginal Distributions

Although the dependence structure captured by the copula does not depend on the choice of the marginal distributions, it is convenient to have suitable marginal distributions. This is why it is important to test if the observed data we have is well represented by a certain distribution. Such tests can be performed in a graphical way, through goodness-of-fit tests or by considering some information criteria measures. One can also compute the coefficient of determination (R^2) to assess the strength of the relationship between the data and the fitted distribution. The higher the value, the better the model in question fits the observed data. If one has a Normal distribution, a test of the normality, such as the Shapiro-Wilk (S-W) test, can be used. Such test can also be performed when the distribution is Lognormal.

Goodness-of-Fit

In order to assess the fit of the marginal distributions to the data, *i.e.*,

$$H_0: X \sim F(x \mid \boldsymbol{\theta}) \quad \text{vs} \quad H_1: X \nsim F(x \mid \boldsymbol{\theta}),$$

where F is some distribution function and $\boldsymbol{\theta}$ is a vector of the parameters, we can perform several tests. For instance, the χ^2 and the Kolmogorov-Smirnov are two of the most used goodness-of-fit tests although the Cramér-von-Mises and the Anderson-Darling tests are commonly used as well. However, one has to take into account if the distribution is fully specified, that is, the true parameters of the underlying population are known, or if the distribution is unspecified, that is, the parameters are unknown and have to be estimated from the data, for example, through MLE. [Stephens, 1986] addresses the unspecified cases of the Normal, the Extreme-Value, the Weibull, the Gamma, the Logistic and the Cauchy distributions. For the scope of this thesis, we are only interested in the Gamma and the Weibull distributions. [Littell *et al*, 1979] also propose goodness-of-fit tests for the two parameter Weibull distribution when the parameters are estimated by maximum likelihood. On the other hand, [Tadikamalla, 1990] proposed modified Kolmogorov-Smirnov type test-statistics for the Gamma, the Erlang-2 and the Inverse Gaussian distributions when the distributions are unspecified.

χ^2-test

The χ^2-test is applied to binned data and, therefore, its value depends on how the data was partitioned. It also requires a sufficiently large sample size so that the approximation to the χ^2 distribution is valid. If $X_1,\ldots,X_n$ is a random sample with distribution function $F(x \mid \theta)$, the χ^2 test statistic for unspecified distributions is given by

$$X^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}, \tag{3.16}$$

where k is the number of cells, O_j the number of observations that lie on the cell j, $E_j = n\int_{B_j} dF(x \mid \hat{\theta})$ is the expected frequency for the bin j, n is the sample size, $\int_{B_j} dF(x \mid \hat{\theta}) = P_{\hat{\theta}}(X_i \text{ falls in } B_j)$, B_j is the cell j and $\hat{\theta}$ is the estimate of θ. Under H_0, $X^2 \overset{a}{\sim} \chi^2_{(k-p-1)}$, where p is the number of parameters to estimate. H_0 is rejected at an approximated significance level α if $X^2 > q_{1-\alpha}$, where $q_{1-\alpha}$ is the $(1-\alpha)$th quantile of a $\chi^2_{(k-p-1)}$ distribution, or for small values of the p-value ($p \leq \alpha$).

[Moore, 1986] enumerates three advantages of the use of χ^2 type statistics. For instance, they are the most practical tests of fit since they are applicable whether the model is specified or unspecified, when we have univariate or multivariate data, which can be discrete, continuous or censored. However, χ^2 tests are usually less powerful than goodness-of-fit tests based on the empirical distribution function (EDF), such as the Kolmogorov-Smirnov, the Anderson-Darling or the Cramér-von-Mises.

Opposed to the previous test statistic, the tests based of the EDF statistics can be used with small samples. EDF statistics measure the discrepancy between the empirical distribution function, (3.7), and the distribution function, $F(x)$, we want to test. In this category lies the well-known Kolmogorov-Smirnov statistic test, D, as well as the Anderson-Darling, A^2, and the Cramér-von-Mises, W^2.

As stated before, when the parameters are unknown and have to be estimated from the data, say by maximum likelihood, one cannot apply the known test statistics, nor use the usual critical values tables. When the unknown parameters are location or scale parameters, [Stephens, 1986] shows that the distributions of EDF statistics do not depend on their true value. However, the exact distributions can be hard to find and one has to perform Monte Carlo studies to find critical points for finite sample sizes. For the scope of this thesis, we will only address the cases of the Lognormal, the Gamma and the Weibull distributions. However, authors such as [Stephens, 1986] and [Tadikamalla, 1990] consider other types of distributions.

Kolmogorov-Smirnov test

Let $X_1,\ldots,X_n$ be a random sample with distribution function $F(x \mid \theta)$, θ the vector of parameters and $x_{(i)}$ the i-th order statistic. The Kolmogorov-Smirnov test statistic is given by

$$D = \sup|F(x \mid \hat{\theta}) - \hat{F}(x)| = \max\{D^+, D^-\}, \tag{3.17}$$

where

$$D^+ = \max_{1\leq i\leq n}\{i/n - F(x_{(i)} \mid \hat{\theta})\} \quad \text{and} \quad D^- = \max_{1\leq i\leq n}\{F(x_{(i)} \mid \hat{\theta}) - (i-1)/n\}. \tag{3.18}$$

For the Gamma and the Weibull distributions, one can use the critical values proposed in [Littell *et al*, 1979] and [Tadikamalla, 1990]. For the Lognormal distribution, we have to take into account that if $X \sim LN(\mu,\sigma^2)$, then $Y = \log(X) \sim N(\mu,\sigma^2)$ and we use the critical values presented by [Lilliefors, 1967].

H_0 is rejected if the test statistic has a higher value than the critical value, at a significance level α.

Anderson-Darling and Cramér-von-Mises tests

In this thesis, we will only use the Anderson-Darling and the Cramér-von-Mises tests for the Gamma and the Weibull distributions and the critical values tables proposed by [Stephens, 1986].

Let again $X_1, \dots, X_n$ be a random sample with distribution function $F(x)$ and $Z_{(i)}$ the i-th order statistic of the sample $Z_1 = F(x_1), \dots, Z_n = F(x_n)$. The Anderson-Darling test statistic is given by

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1)[\log(Z_{(i)}) + \log(1 - Z_{(n+1-i)})], \tag{3.19}$$

and the Cramér-von-Mises test statistic by

$$W^2 = \sum_{i=1}^{n} \left(Z_{(i)} - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}. \tag{3.20}$$

For the unspecified Weibull distribution, [Stephens, 1986] constructed critical values tables with the following modification on the test statistics

$$A^{*2} = A^2 \left(1 + \frac{1}{5\sqrt{n}} \right) \quad W^{*2} = W^2 \left(1 + \frac{1}{5\sqrt{n}} \right). \tag{3.21}$$

For the unspecified Gamma distribution, [Stephens, 1986] proposed a more complicated modification. Considering α to be the shape parameter and β to be the rate parameter, the following procedure has to be carried out to perform both tests:

1. Estimate α by solving for $\hat{\alpha}$

$$\frac{1}{n} \sum_{i=1}^{n} \log(X_i) - \log(\overline{X}) = \psi(\alpha) - \log(\alpha), \tag{3.22}$$

   where $\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ is the digamma function, and estimate β by $\hat{\beta} = \frac{\hat{\alpha}}{\overline{X}}$.

2. Calculate (3.19) and (3.20) with $Z_{(i)} = \gamma(x_{(i)} \mid \hat{\alpha}, \hat{\beta})$, where $\gamma(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-y\beta} \, \mathrm{d}y$.

As usual, H_0 is rejected if the value of the test statistic is higher than the critical value, at a significance level α.

Shapiro-Wilk test

Another possibility for testing if the observed data can be modelled by a Lognormal distribution was presented by [González-Estrada and Villaseñor, 2018] and it follows the same line of argument of the one proposed by [Lilliefors, 1967]: if $X \sim LN(\mu, \sigma^2)$, then $Y = \log(X) \sim N(\mu, \sigma^2)$. Therefore, it is plausible to test the hypothesis of normality of Y. H_0 will be rejected for low values of the p-value.

Akaike's Information Criterion

Contrary to the previous tests, AIC does not provide information of the general fit of the model. However, it can be a useful tool for comparing different models, for instance, if a certain distribution describes better the data than other. It was first introduced in [Akaike, 1974] and its expression relies on the concept of log-likelihood. Let $X_1, \dots, X_n$ be a random sample with distribution function $F(X)$. The log-likelihood for an univariate continuous distribution is defined as

$$l(\boldsymbol{\theta} \mid \boldsymbol{x}) = \sum_{i=1}^{n} \log(f(x_i \mid \boldsymbol{\theta})). \tag{3.23}$$

The AIC is then given by
$$AIC = 2k - 2l(\theta \mid \boldsymbol{x}), \tag{3.24}$$
where k is the number of components of the vector of parameters, $\boldsymbol{\theta}$. The best model will be the one with the smallest AIC value.

$\boldsymbol{R^2}$

The last topic we will cover is the one regarding the R^2. Just like the AIC, it does not give any information of the general fit of a model. It is only a measure of the strength of the relationship between the data and the fitted distribution.

In order to assess if a particular model provides a plausible fit to the distribution of the variable of the observed data, one can use graphical tools such as the well-known Quantile-Quantile plots, QQ-Plots. [Beirlant *et al*, 2004] provide a table with QQ-plot coordinates for some distributions, such as for the Lognormal, Exponential and the Weibull distributions. For other distributions, which are not location--scale models, such as the Gamma or the Burr, they suggest transforming the data into the exponential case. Let $X_1, \dots, X_n$ be a random sample of X with a distribution function $F(x \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of parameters and $x_{(i)}$ represent the i-th order statistic.

$$E_{(i)} = -\log(1 - F(x_{(i)} \mid \boldsymbol{\theta})) \tag{3.25}$$

is the i-the order statistic associated with a random sample of size n from the standard exponential distribution, $E \sim Exp(1)$. One can then construct an exponential QQ-Plot, which compares the empirical quantiles with the exponential quantiles, with the following coordinates

$$(-\log(1 - p_{(i)}), -\log(1 - F(x_{(i)} \mid \hat{\boldsymbol{\theta}})), \tag{3.26}$$

where $p_{(i)} = \frac{i}{n+1}$, $i = 1, 2, \dots, n$. After plotting the coordinates, one can perform a linear regression on the resulting QQ-plot, and then measure the global fit of the distribution through the correlation coefficient, or through the R^2; see [Beirlant *et al*, 2004]. The higher value, the better the model fits to the variable in hand.

3.2.2 Copula

Information Criteria

The information criteria techniques are useful to provide us the loss incurred of using a certain parametric model M instead of the true one. Therefore, the smaller values of the criteria, the lower the loss and, thus, the better the model. The information criteria are also invariant to monotone increasing transformations of the marginal distributions, which is a very appealing characteristic when modelling with copulas; see [Dos Santos Silva and Lopes, 2008].

Let $L(\boldsymbol{\theta} \mid \boldsymbol{z})$ be the likelihood function, (3.14). The deviance of M is given by

$$D(\boldsymbol{\theta}) = -2\log(L(\boldsymbol{\theta} \mid \boldsymbol{z})) \tag{3.27}$$

where $\boldsymbol{\theta}$ is the k-dimensional parameter vector. We can see that the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}_{MLE}$, will give the smallest value of deviance.

Similarly to (3.24), we can define the Akaike's Information Criterion (AIC) of model M by

$$AIC(M) = D(\hat{\boldsymbol{\theta}}_{MLE}) + 2k. \tag{3.28}$$

This criterion is mostly determined by the smallest value of the deviance but also offers a penalisation for larger models. The lower the dimension of vector $\boldsymbol{\theta}$, the better M is.

[Schwarz, 1978] suggested the Bayesian Information Criterion (BIC) another information criterion which is defined by
$$BIC(M) = D(\hat{\boldsymbol{\theta}}_{MLE}) + k\log(n), \tag{3.29}$$

where n is the sample size. We can see that this criterion penalises even more the over-parameterised models. According to the BIC, smaller samples provide better models.

These two criteria allow us to compare several models which may or may not have different parametric dimensions. This is an important advantage, because one can have two models with different vectors of parameters and still be able to select one over the other.

In a Bayesian framework, none of the above criteria is particularly acceptable to compare and select models. As we can see from (3.28) and (3.29), they are based on the value of the MLE for the vector $\boldsymbol{\theta}$, which may or may not be available explicitly in Bayesian estimation. They also do not consider the posterior distribution. [Spiegelhalter *et al*, 2002] shows that the inclusion of a prior distribution may reduce the effective dimensionality of the model, since it will cause a dependence between its parameters and it seems reasonable to have a measure which depends on both the prior information and the observed data. Therefore, they propose a new information criterion called Deviance Information Criterion, DIC, which is defined as

$$DIC(M) = D(\overline{\boldsymbol{\theta}}) + 2(\overline{D} - D(\overline{\boldsymbol{\theta}})) = 2\overline{D} - D(\overline{\boldsymbol{\theta}}), \tag{3.30}$$

$D(\overline{\boldsymbol{\theta}})$ is the deviance at the posterior mean and $\overline{D}$ is the posterior mean of the deviance $D(\boldsymbol{\theta})$. The first term of the DIC characterises the fit of the model while the second penalises the over-parameterisation; [Shemyakin and Kniazev, 2017]. An advantage of this criteria measure is its application to MCMC procedures since it is possible to estimate $D(\overline{\boldsymbol{\theta}})$ and $\overline{D}$ directly from the generated sample. Again, the lower are the values of DIC, the better M is.

Goodness-of-Fit

As stated above, goodness-of-fit tests are more difficult to perform in the multivariate case than in the univariate case, which has a variety of tests available. When we are in the bivariate case, these tests are not trivial and require more effort. However, this problem has been addressed by some authors. [Wang and Wells, 2000] and [Rosenblatt, 1952] propose goodness-of-fit tests basing the null hypothesis on integral transforms of the data, while [Genest *et al*, 2009] propose tests where the marginal distributions are considered to be unknown. Later, [Berg, 2009] proposes an extension to the parametric case where the marginal distributions are known.

Let $\boldsymbol{X} = (X,Y)$ be a random continuous vector with joint distribution H and margins $F(x)$ and $G(y)$. Similarly to the univariate case, we want to test

$$H_0 : C \in C_0 \quad vs \quad H_1 : C \notin C_0$$

where $C_0 = \{C_\theta : \theta \in \Theta\}$ and $\Theta \in \mathbb{R}^2$ is the domain of the vector of parameters θ.

[Genest *et al*, 2009] shows that modelling the margins by parametric families is not viable if we are testing H_0 since we are checking if the dependence of a bivariate distribution is well-represented by the family of copulas C_0. Therefore, for testing the null hypothesis, we base the inference on the pseudo-observations defined in (3.10).

The tests proposed in [Genest *et al*, 2009] are based on the empirical copula, (3.11), and consist on a comparison between the empirical copula, C_n, and an estimation of the copula under H_0, $C_{\hat{\theta}}$, where $\hat{\theta}$ is an estimate of θ derived from the pseudo-observations. Moreover, they are based on the empirical process

$$\mathbb{C}_n = \sqrt{n}\left(C_n - C_{\hat{\theta}}\right). \tag{3.31}$$

They defined the rank-based versions of Cramér-von-Mises and Kolmogorov-Smirnov statistics, respectively, S_n and T_n, as

$$S_n = \int_{I^2} \mathbb{C}_n(u,v)^2 \,\mathrm{d}C_n(u,v) \quad \text{and} \quad T_n = \sup_{u,v \in I^2} |\mathbb{C}_n(u,v)|. \tag{3.32}$$

These statistics do not follow known distributions and thus the p-values have to be obtained via bootstrapping.

On the other hand, the tests proposed in [Wang and Wells, 2000] are based on Kendall's transform of the data, $\boldsymbol{X} \mapsto W = H(\boldsymbol{X}) = C(U,V)$ with $U = F(X)$ and $V = G(Y)$. Let K be the univariate distribution function of W and $W_1 = C_n(\boldsymbol{Z}_1), \dots, W_n = C_n(\boldsymbol{Z}_n)$, where $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ are the vectors of the pseudo-observations defined in (3.10). K can be estimated by the empirical distribution

$$K_n(w) = \frac{1}{n}\sum_{i=1}^{n} I_{[0,1]}(W_i \leq w). \tag{3.33}$$

As before, the tests are based on the empirical process

$$\mathbb{K}_n = \sqrt{n}\left(K_n - K_{\hat{\theta}}\right), \tag{3.34}$$

although they are not generally consistent. The rank-based versions of Cramér-von-Mises and Kolmogorov-Smirnov $S_n^{(K)}$ and $T_n^{(K)}$, respectively, are given by

$$S_n^{(K)} = \int_0^1 \mathbb{K}_n(w)^2 \, \mathrm{d}K_{\hat{\theta}}(w) \quad \text{and} \quad T_n^{(K)} = \sup_{w\in[0,1]} |\mathbb{K}_n(w)|. \tag{3.35}$$

Once more, the p-values have to be obtained via simulation.

In turn, [Rosenblatt, 1952] proposed a probability integral transformation which allows to decompose a random vector with a given distribution into mutually independent components uniformly distributed on the unit interval.

Definition 3.2.1. The *Rosenblatt's probability integral transform* of a copula C is the mapping $R: (0,1)^2 \to (0,1)^2$, which assigns to every $\boldsymbol{u} = (u,v) \in (0,1)^2$, another vector $R(\boldsymbol{u}) = (e_1, e_2)$ with $e_1 = u$ and

$$e_2 = \frac{\dfrac{\partial C(u,v)}{\partial u}}{\dfrac{\partial C(u,1)}{\partial u}}. \tag{3.36}$$

Note that $\boldsymbol{U}$ is distributed as C if and only if the distribution of $R(\boldsymbol{U})$ is the bivariate product copula, $\Pi(e_1,e_2) = e_1 \times e_2$, $e_1, e_2 \in [0,1]$.

Let $\boldsymbol{E}_1 = R_{\hat{\theta}}(\boldsymbol{U}_1), \dots, \boldsymbol{E}_n = R_{\hat{\theta}}(\boldsymbol{U}_n)$ be pseudo-observations, which can be interpreted as a sample from the product copula Π. The empirical distribution is

$$D_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} I_{[0,1]^2}(\boldsymbol{E}_i \leq \boldsymbol{u}) \tag{3.37}$$

and, under H_0, it should be close to the product copula. Two Cramér-von-Mises statistics, which only differ in the integration measure, are considered

$$S_n^{(C)} = n\int_{[0,1]^2} [D_n(\boldsymbol{u}) - \Pi(\boldsymbol{u})]^2 \, \mathrm{d}D_n(\boldsymbol{u}) = \sum_{i=1}^{n} [D_n(\boldsymbol{E}_i) - \Pi(\boldsymbol{E}_i)]^2 \tag{3.38}$$

and

$$S_n^{(B)} = n\int_{[0,1]^2} [D_n(\boldsymbol{u}) - \Pi(\boldsymbol{u})]^2 \, \mathrm{d}\boldsymbol{u} = \frac{n}{9} - \frac{1}{2}\sum_{i=1}^{n}\prod_{t=1}^{2}(1 - E_{it}^2) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\prod_{t=1}^{2}\max\{1 - E_{it}, E_{jt}\}. \tag{3.39}$$

[Genest *et al*, 2009] performed a Monte Carlo experiment and, although none of the tests are preferable to any of the others, they concluded that the tests based on the Kolmogorov-Smirnov seem to be much less powerful than the others, and they do not advise using these type of tests. To their knowledge, statistics S_n (3.32) and $S_n^{(B)}$ (3.39) yield the best goodness-of-fit tests for copula models, with the latter being more consistent. They also recommend the use of statistics $S_n^{(K)}$ (3.35) and $S_n^{(C)}$ (3.38). Moreover, on the presence of Archimedean copulas $S_n^{(K)}$ is especially convenient, since K is available on a closed form. Based on the results of their simulation, we have $S_n^{(B)} \succ S_n \succ S_n^{(K)} \succ S_n^{(C)} \succ T_n \succ T_n^{(K)}$. [(a)]

[(a)] $a \succ b$ means that a succeeds b.

4 | Procedure and Results

The implementations of this chapter were done in R. For most of the distributions, the package *fitdistrplus* proposed by [Delignette-Muller and Dutang, 2015] was used to fit the marginals and estimate their parameters by maximum likelihood. The packages *actuar* and *ExtDist* were used to fit the Burr distribution. The goodness-of-fit tests for the marginal distributions were computed with the function *gofstat* of the *fitdistrplus* package and using the packages *goft*; see [González-Estrada and Villaseñor, 2018], and *KScorrect*; see [Novack-Gottshall and Wang, 2016]. The copulas were implemented using the *copula*; see [Yan, 2007] and the *VineCopula*; see [Schepsmeier *et al*, 2018] packages and their fit was assessed using the package *gofCopula*; see [Okhrin *et al*, 2018]. Regarding the Bayesian inference, we used the package *runjags*; see [Denwood, 2016] and the JAGS software. However, a script of the implementation of the Bayesian inference for copula modelling using WinBUGS can be found on Appendix C.

4.1 Wind Speed Data

This thesis uses copulas to analyse the dependence between daily maximum wind speeds, X, measured in km/h, observed in 40 stations spread out in the continental part of Portugal from 2000 to 2012, and simulated daily maximum wind speeds, Y, produced by a simulator, at a regular grid of $81km^2$ grid cell size. In reality, there are 117 stations on the meteorological stations network of the Instituto Português do Mar e da Atmosfera, IPMA. However, due to the fact that the observed data has an extremely high proportion of missing observations, which reaches 90% in some stations, we have only considered the ones with less than 30% of NAs, resulting in 40 out of 117. Although the data was initially from 1997 to 2013, from 1997 until 2000 large periods without any record were observed and 2013 only had observations until the month of February. Therefore, we have decided to analyse just the years, which are complete or which have very few NAs, that is, from 2000 to 2012, see Tab. 4.1 and Tab. 4.2. The missing values were removed from the observed data as well as the wind speeds equal to 0, since theoretically it is impossible to have no wind and we suspect they were either errors of the meteorological station or missing values. The observed data was provided by the IPMA and the simulated data by the Instituto Dom Luiz of the University of Lisbon.

Due to size and space constraints, and in order not to be exhaustive, we will only present the results for 9 stations: Aveiro, Bragança, Castelo Branco, Coruche, Estremoz, Lisboa, Monção, Sines and Vila do Bispo, while the remaining ones can be found in Appendix B. We have also tried to choose stations which cover and capture each region of the country. We have two meteorological stations in Lisboa, one located in Rua da Escola Politécnica, in Rato, which we will refer to as Station 1 (S1), and the other located in Avenida Gago Coutinho, near Portela, which we will denote by Station 2 (S2).

Tab. 4.1: Number of observations recorded per year in the 9 selected meteorological stations, from 1997 to 2005.

Station	1997	1998	1999	2000*	2001	2002	2003	2004*	2005
Aveiro	57	176	338	364	360	365	357	366	365
Bragança	–	214	365	352	363	365	365	366	345
Castelo Branco	–	216	363	366	365	364	365	366	365
Coruche	187	288	266	363	362	344	355	364	361
Estremoz	214	355	362	361	362	359	362	364	365
Lisboa S1	–	–	212	366	365	365	365	366	365
Monção	7	206	300	339	339	355	343	355	331
Sines	184	334	265	366	365	365	365	366	365
Vila do Bispo	197	297	356	349	365	365	365	366	365

* denotes the leap years.

Tab. 4.2: Number of observations recorded per year in the 9 selected meteorological stations, from 2006 to 2013.

Station	2006	2007	2008*	2009	2010	2011	2012*	2013
Aveiro	348	365	359	347	332	360	363	59
Bragança	365	364	366	363	365	365	365	59
Castelo Branco	365	362	366	365	341	365	366	59
Coruche	255	341	351	362	352	364	365	57
Estremoz	364	365	363	357	354	285	362	–
Lisboa S1	182	362	366	363	362	365	365	59
Monção	353	360	325	362	356	363	358	58
Sines	365	365	363	365	365	365	289	59
Vila do Bispo	365	363	366	365	365	356	299	59

* denotes the leap years.

4.2 Data Preprocessing

If we think about the wind in Portugal, which has a mediterranean climate, we can deduce that it has a seasonal pattern. For instance, one is expecting higher values of wind speed during the Winter season and lower values during Summer, which reflects the non-stationary behaviour of the wind speed throughout the years; see [Naveau *et al*, 2016]. For this motive, we have decided to divide the data set into 4 parts, each of which represents one season. The seasonal cycle of the wind of the 9 stations considered is presented in Fig. 4.1. We can see that, for some stations like Castelo Branco or Lisboa S1, the observed wind is mostly similar between the first and third quartile in a set of 2 seasons: Spring + Summer and Autumn + Winter. However, the wind in Aveiro or in Vila do Bispo appears to have a different behaviour for each season. Therefore, we think that in the overall the choice of dividing the data into seasons was appropriate. Tables 4.3 and 4.4 summarise the informations above. One can note that, in some cases, especially in Spring, the simulated wind appears to be shifted to the right relatively to the observed wind. This can be a problem, since it means that the simulator does not reflect the real wind and does not produce accurate values.

We have also retained every 5 observations in every season and station in order to overcome the short-term dependence that exists within each time series. In Fig. 4.2, we have the autocorrelation function plots for the station of Castelo Branco. The ACF plots for the other 8 meteorological stations can be found in Appendix B. Although some situations still show autocorrelation (for instance the observed wind in Castelo Branco's Summer), most of the cases no longer reveal any significant dependence within each series.

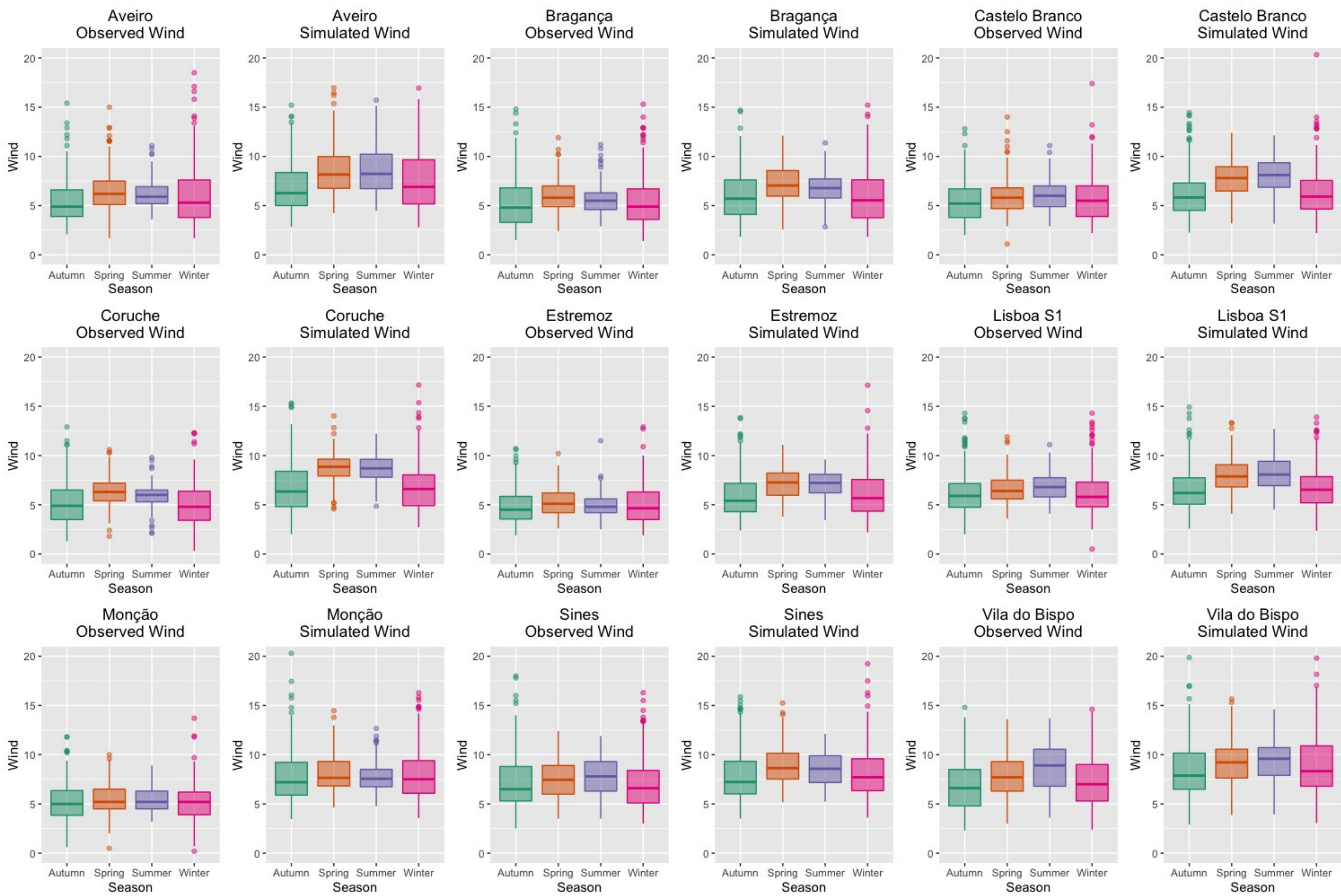


Fig. 4.1: Boxplots of the observed and simulated wind speed data for each season of the selected 9 stations: Aveiro, Bragança, Castelo Branco, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo.

Tab. 4.3: Summary of the values of the observed, X, and simulated, Y, winds, measured in km/h, in the selected 9 stations in the seasons of Autumn and Winter.

	Autumn							**Winter**						
	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
X_{Aveiro}	234	2.1	3.9	4.9	5.468	6.6	15.4	231	1.7	3.8	5.3	6.13	7.6	18.5
Y_{Aveiro}		2.833	5.006	6.270	6.930	8.358	15.199		2.8	5.167	6.9	7.521	9.653	16.937
$X_{Braganca}$	235	1.5	3.3	4.8	5.351	6.8	14.8	232	1.4	3.6	4.9	5.478	6.7	15.3
$Y_{Braganca}$		1.836	4.105	5.709	6.011	7.604	14.683		1.825	3.776	5.545	5.912	7.612	15.198
$X_{CasteloBranco}$	237	2	3.8	5.2	4.453	6.7	12.8	233	2.2	3.9	5.5	5.73	7	17.4
$Y_{CasteloBranco}$		2.255	4.519	5.823	6.217	7.298	14.445		2.202	4.663	5.917	6.354	7.556	20.331
$X_{Coruche}$	215	1.3	3.5	4.9	5.232	6.5	12.9	222	0.3	3.425	4.8	5.083	6.375	12.3
$Y_{Coruche}$		2.029	4.815	6.347	6.859	8.401	15.323		2.711	4.917	6.602	6.806	8.036	17.161
$X_{Estremoz}$	235	1.9	3.55	4.5	4.911	5.85	10.7	212	1.9	3.5	4.65	4.974	6.3	12.9
$Y_{Estremoz}$		2.384	4.299	5.408	5.965	7.172	13.837		2.204	4.35	5.683	6.128	7.563	17.14
$X_{LisboaS1}$	219	2	4.75	5.9	6.261	7.15	14.3	229	0.5	4.8	5.8	6.233	7.3	14.3
$Y_{LisboaS1}$		2.565	5.06	6.2	6.582	7.744	14.918		2.354	5.199	6.532	6.826	7.852	13.895
X_{Moncao}	227	0.6	3.85	5	5.149	6.35	11.8	214	0.2	3.9	5.2	5.239	6.2	13.7
Y_{Moncao}		3.465	5.897	7.201	7.833	9.231	20.293		3.562	6.09	7.507	8.047	9.394	16.274
X_{Sines}	237	2.5	5.3	6.5	7.235	8.8	18	233	3	5.1	6.6	7.189	8.4	16.3
Y_{Sines}		3.522	6.022	7.228	7.888	9.332	15.882		3.603	6.341	7.706	8.265	9.584	19.23
$X_{VilaBispo}$	235	2.3	4.8	6.6	6.993	8.5	14.8	217	2.4	5.3	7	7.262	9	14.6
$Y_{VilaBispo}$		2.903	6.489	7.876	8.479	10.16	19.865		3.053	6.803	8.325	8.983	10.891	19.803

Tab. 4.4: Summary of the values of the observed, X, and simulated, Y, winds, measured in km/h, in the selected 9 stations in the seasons of Spring and Summer.

	Spring							Summer						
	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
X_{Aveiro}	231	1.7	5.1	6.2	6.525	7.5	15	236	3.6	5.2	5.9	6.173	6.925	11.1
Y_{Aveiro}		4.242	6.762	8.158	8.494	9.984	16.96		4.467	6.724	8.225	8.559	10.218	15.708
$X_{Braganca}$	237	2.4	4.9	5.8	6.05	7	11.9	239	2.9	4.6	5.5	5.577	6.3	11.2
$Y_{Braganca}$		2.577	5.963	7.041	7.263	8.557	12.099		2.848	5.778	6.781	6.761	7.713	11.374
$X_{CasteloBranco}$	240	1.1	4.7	5.8	5.936	6.8	14	237	2.9	4.9	6	6.03	7	11.1
$Y_{CasteloBranco}$		3.199	6.485	7.798	7.855	8.952	12.389		3.168	6.866	8.106	8.046	9.349	12.126
$X_{Coruche}$	239	1.8	5.4	6.3	6.347	7.2	10.6	233	2.1	5.3	6	5.945	6.5	9.8
$Y_{Coruche}$		4.574	7.921	8.854	8.78	9.63	14.025		4.865	7.798	8.699	8.722	9.613	12.218
$X_{Estremoz}$	241	2.6	4.2	5.1	5.212	6.2	10.2	239	2.5	4.2	4.8	5.015	5.6	11.5
$Y_{Estremoz}$		3.784	5.965	7.268	7.218	8.227	11.087		3.433	6.213	7.221	7.07	8.092	9.598
$X_{LisboaS1}$	242	3.6	5.6	6.4	6.578	7.5	11.9	223	4.1	5.8	6.8	6.873	7.75	11.1
$Y_{LisboaS1}$		4.089	6.814	7.88	7.979	9.061	13.338		4.507	6.943	8.065	8.242	9.416	12.711
X_{Moncao}	236	0.5	4.5	5.2	5.542	6.5	10	233	3.2	4.5	5.2	5.488	6.3	8.9
Y_{Moncao}		4.63	6.825	7.646	8.124	9.313	14.463		4.759	6.747	7.557	7.745	8.513	12.663
X_{Sines}	226	3.5	6	7.45	7.586	8.9	12.4	240	3.5	6.3	7.8	7.804	9.3	11.9
Y_{Sines}		5.161	7.527	8.634	8.907	10.146	15.245		5.229	7.175	8.569	8.564	9.897	12.12
$X_{VilaBispo}$	242	3.3	6.3	7.7	7.91	9.3	13.6	239	3.6	6.8	8.9	8.683	10.55	13.7
$Y_{VilaBispo}$		3.868	7.634	9.219	9.16	10.552	15.655		3.943	7.899	9.588	9.406	10.713	14.61

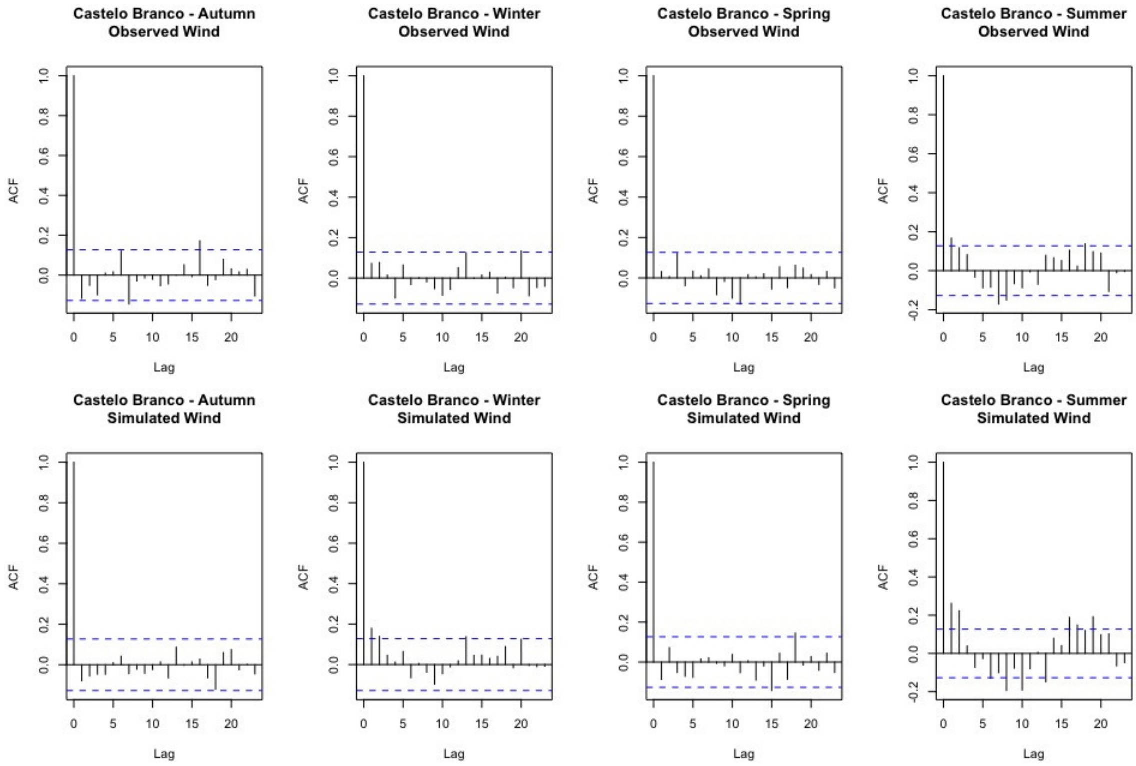


Fig. 4.2: ACF of the observed and simulated wind speed data for each season of Castelo Branco's station.

4.3 Modelling Wind Speed Data

In the literature, the Weibull distribution is the most commonly used to model the wind speed; see [Mert and Karakus, 2015], [Pobocikova *et al*, 2017], [Shepherd, 1978] and [Harris and Cook, 2014]. However, some authors, such as [Çelik and Yilmaz, 2008], remark that one should not use the Weibull distribution to model the wind speed without a previous statistical analysis. Additionally, other distributions such as Lognormal and Gamma can be used to model wind type data; see [Pobocikova *et al*, 2017] and [Harris and Cook, 2014]. It is observed in [Mert and Karakus, 2015] that the Weibull distribution does not model properly the wind in regions where it is common to have low winds. Therefore, they proposed to fit a 4-parameter Burr distribution, which they claim to be more appealing to fit wind speed data. For these reasons, in this thesis we will consider four distributions: the Weibull, the Lognormal, the Gamma and the 3-parameter Burr. To assess their fit, we will use the AIC, the R^2 and the goodness-of-fit tests mentioned in Chapter 3.

4.4 Classical Approach

We will first fit univariate distributions to the observed wind, X, and to the simulated wind, Y. Then, after assessing their fit to the data, we will study the dependence between the two variables and find the copula model which captures it best.

4.4.1 Marginal Distributions

As stated before, we will consider 4 possible distributions which are commonly used to model wind speed data: the Lognormal, the Gamma, the Weibull and the 3-parameter Burr distributions. The parameterisations used are presented in Appendix A. To illustrate all the steps of each procedure, and due to space limitations, we decided to present a complete analysis for a meteorological station. We selected Castelo Branco. However, we will just show the plots for the Winter season of Castelo Branco, while the others can be found on Appendix B. Later, we will present the results for the remaining 8 stations and their seasons.

Tables 4.5 and 4.6 sum up the results all the goodness-of-fit performed to the observed and to the simulated data sets. As mentioned before, we test: $H_0 : X \sim F(x \mid \boldsymbol{\theta})$ vs $H_1 : X \not\sim F(x \mid \boldsymbol{\theta})$, where F is a theoretical distribution function and $\boldsymbol{\theta}$ is the vector of parameters to be estimated using the data. In our case, F is either Lognormal, Gamma, Weibull or Burr. For example for the $LN(\mu, \sigma)$, we have: $H_0 : X \sim LN(\mu, \sigma)$ vs $H_1 : X \not\sim LN(\mu, \sigma)$, where the model parameters should be estimated using the available data. Similar hypotheses are defined for the other models. The Lognormal has the lowest AIC between the 4 models for both data sets. Therefore, it should be preferred to the others. R^2 is also the largest of the 4 models.

In the case of the observed data, at the usual significance levels, Lognormal, Gamma and Burr distributions are not rejected by the χ^2-test; see Tab. 4.5. However, taking into account the drawbacks of this test mentioned in Section 3.2.1, one should perform more suitable tests. The Kolmogorov-Smirnov, the Anderson-Darling and the Cramér-von-Mises tests reject the Gamma and the Weibull distributions, while the Lognormal is not rejected neither by the KS test nor by the Shapiro-Wilk test, at a significant level of 5%. For these reasons, one should choose the Lognormal as the distribution which best fits our data. Fig. 4.3 presents several graphical assessments of the fit of the observed (top) and simulated (bottom) data to the Lognormal, Gamma, Weibull and Burr models. Fig. 4.4 shows the QQ-plots for both data sets along with the corresponding confidence intervals (95%) and Fig. 4.5 exhibits the parametric models with the kernel density estimates. Nonetheless, if we look at the upper tail on Fig. 4.4, Gamma seems to fit better than Lognormal. The fit in the upper tail is extremely important, since this tail represents the higher values of wind and, as we stated before, it is the strong wind which causes greater damage. In spite of that, we consider that in the overall Lognormal should be the chosen model.

Tab. 4.5: Goodness-of-fit tests' results for the marginal distributions fitted to the **observed wind** of Castelo Branco's Winter season data set, values of AIC and R^2.

Distribution	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Lognormal	0.1278	H_0 not rejected	–	–	H_0 not rejected	**1011.0699**	0.9909
Gamma	0.0502	H_0 rejected	H_0 rejected	H_0 rejected	–	1017.6566	0.9682
Weibull	0.0004	H_0 rejected	H_0 rejected	H_0 rejected	–	1045.6639	0.9261
Burr	0.0174	–	–	–	–	1024.5307	0.9905

– These goodness-of-fit tests are not implemented for the distributions in question.

As for the simulated data, Lognormal, Gamma and Burr distributions are clearly not rejected by the χ^2-test. Weibull is rejected by all the tests performed, while Lognormal and Gamma are not rejected by any; see Tab. 4.6. Additionally, if we look at Fig. 4.3, Fig. 4.4 and Fig. 4.5, Lognormal clearly a fits well the simulated data, including also the upper tail.

Tab. 4.6: Goodness-of-fit tests' results for the marginal distributions fitted to the **simulated wind** of Castelo Branco's Winter season data set, values of AIC and R^2.

Distribution	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Lognormal	0.6316	H_0 not rejected	–	–	H_0 not rejected	**1024.9423**	0.9960
Gamma	0.4655	H_0 not rejected	H_0 not rejected	H_0 not rejected	–	1033.4992	0.9255
Weibull	0.0008	H_0 rejected	H_0 rejected	H_0 rejected	–	1073.8739	0.9394
Burr	0.4363	–	–	–	–	1029.3653	0.9945

– These goodness-of-fit tests are not implemented for the distributions in question.

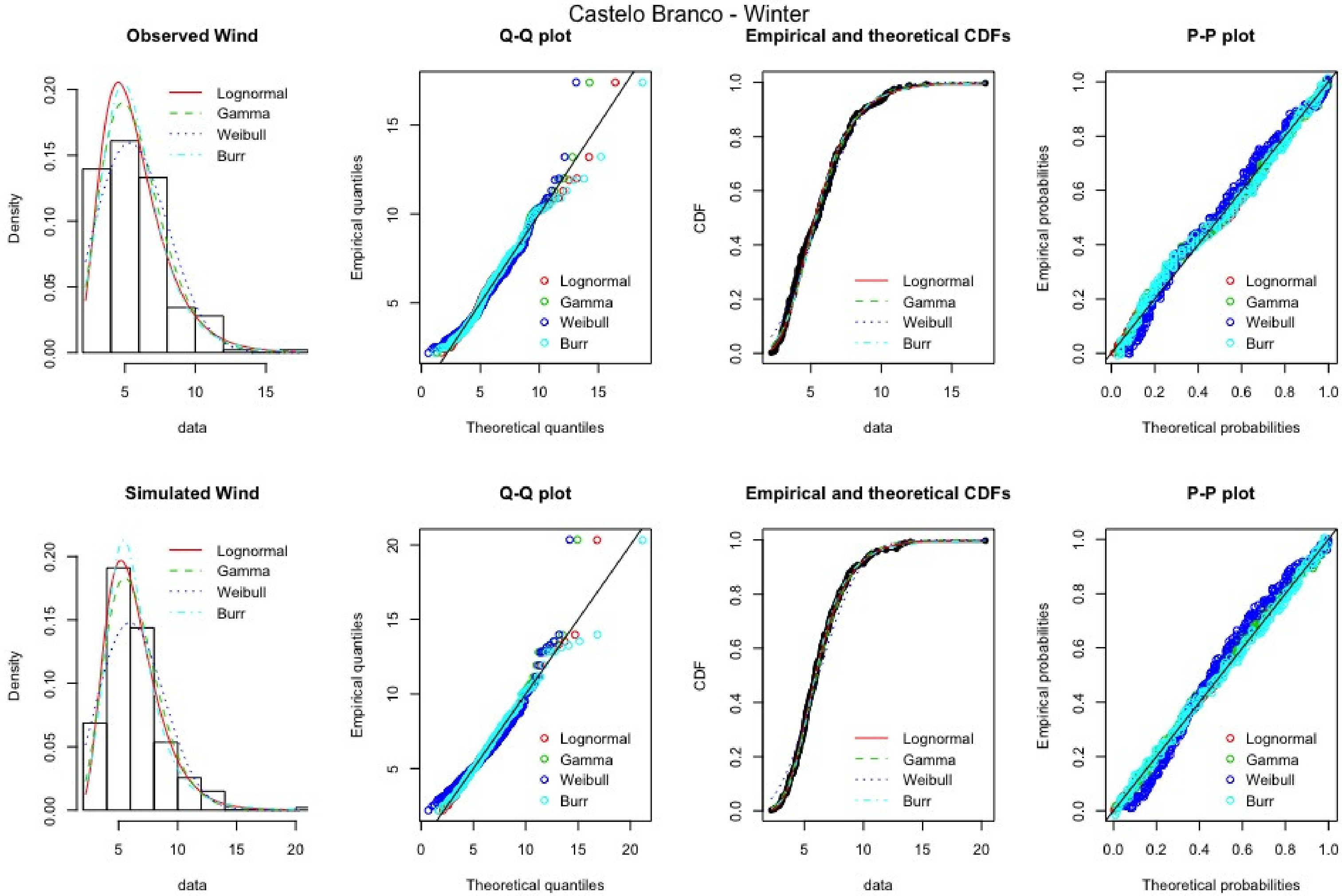


Fig. 4.3: Comparison between the 4 distributions fitted to Winter of Castelo Branco.

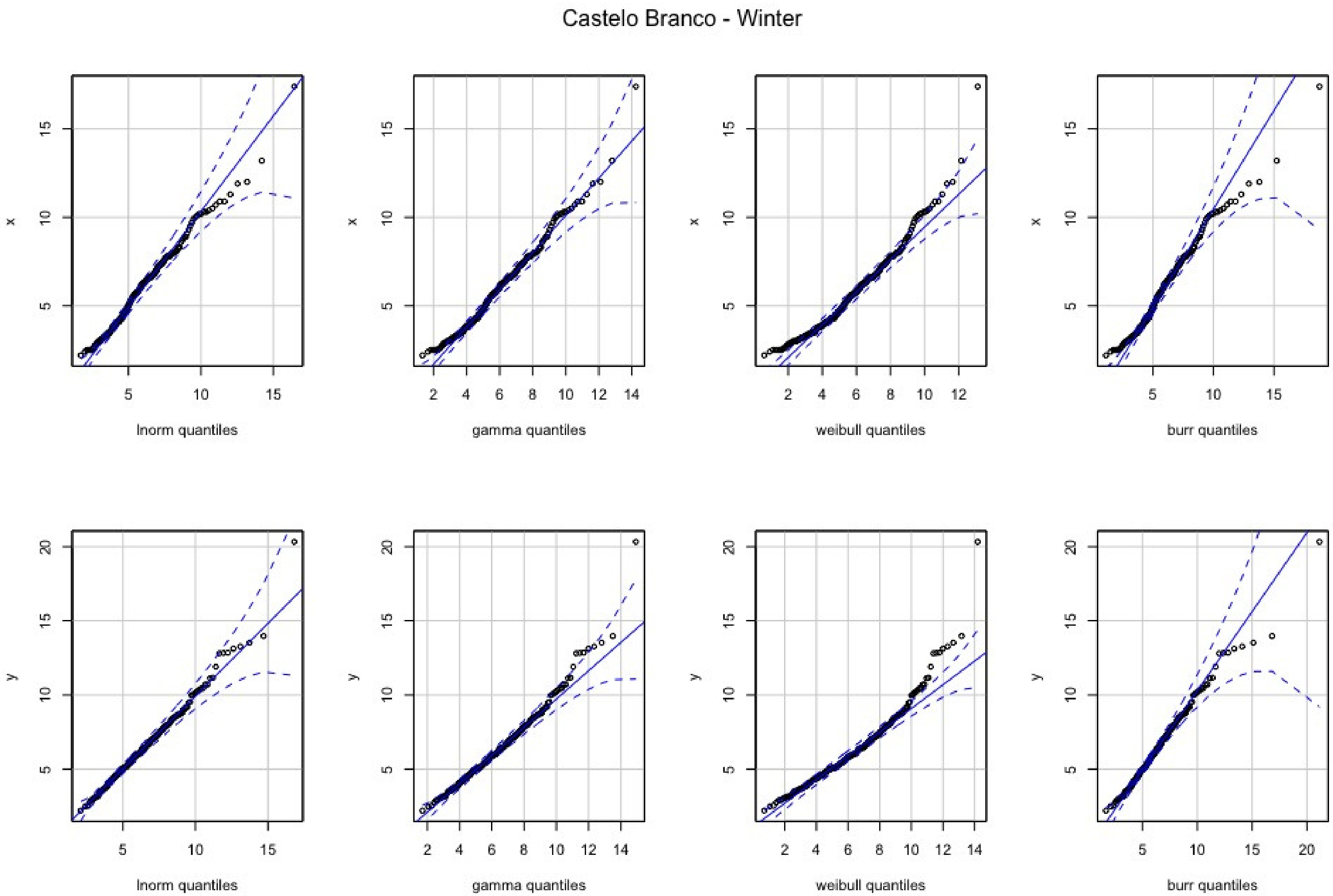


Fig. 4.4: qqPlots of the 4 distributions fitted to Winter of Castelo Branco.

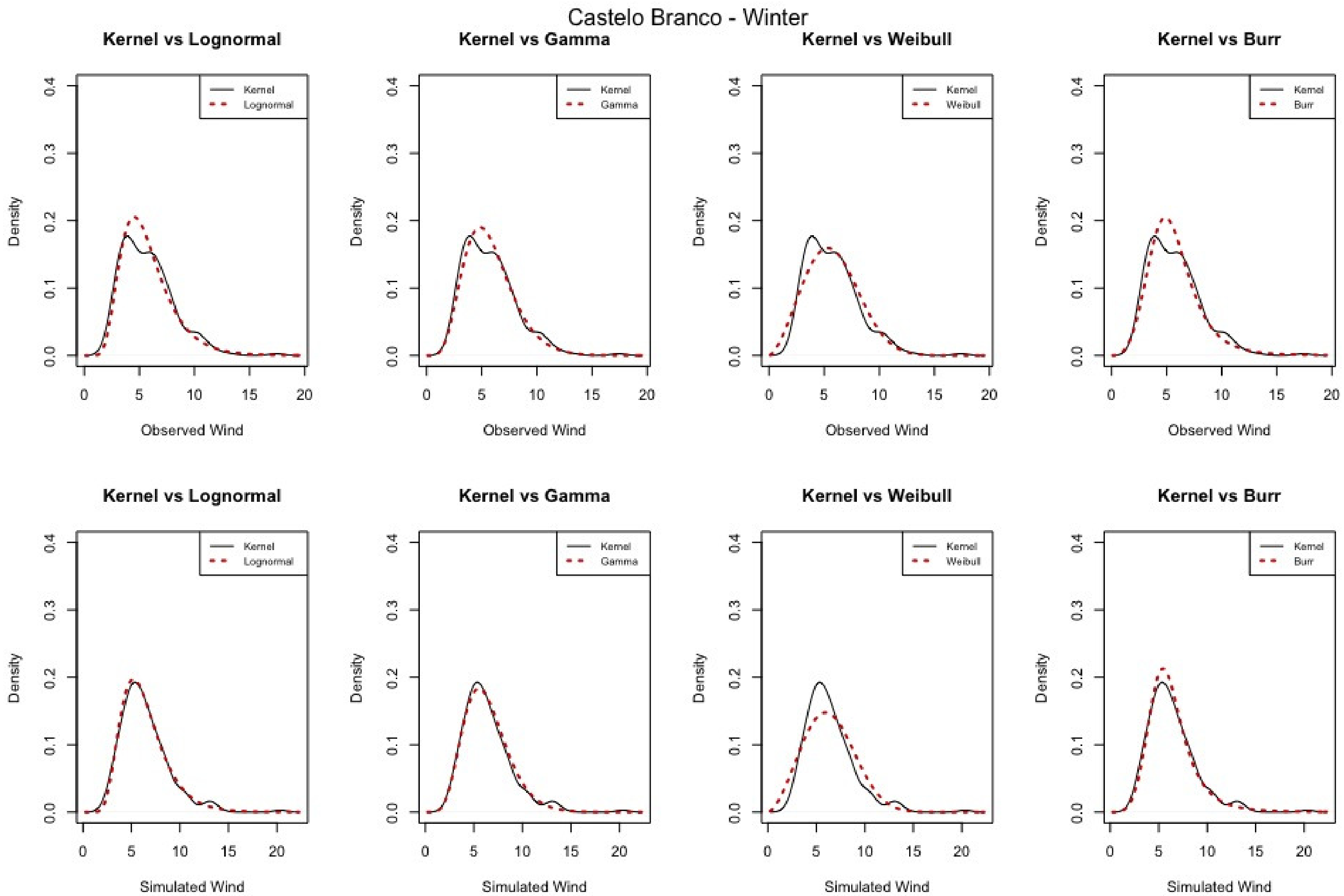


Fig. 4.5: Kernel vs Fitted plots of the 4 distributions fitted to Winter of Castelo Branco.

Thus, we have

$$X_W \sim LN(\mu_W^X, \sigma_W^X) \quad \text{and} \quad Y_W \sim LN(\mu_W^Y, \sigma_W^Y), \tag{4.1}$$

where X_W and Y_W are the observed and simulated wind speed data for the Winter season, respectively.

The MLE estimates and the 95% confidence intervals for $\mu_W^X, \sigma_W^X, \mu_W^Y, \sigma_W^Y$ are presented in Tab. 4.7. The 95% confidence intervals were computed via bootstrap, using the function *bootdist* of the *fitdistrplus* package in R.

Tab. 4.7: Fitted distributions to the observed, X, and simulated, Y, winds of Castelo Branco's station in Winter. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

		$\hat{\theta}_{MLE}$	CI (95%)			$\hat{\theta}_{MLE}$	CI (95%)
X_W				Y_W			
LN	$\hat{\mu}_W^X$	1.6670	(1.6155,1.7175)	LN	$\hat{\mu}_W^Y$	1.7819	(1.7358,1.8271)
	$\hat{\sigma}_W^X$	0.3966	(0.3586,0.4310)		$\hat{\sigma}_W^Y$	0.3642	(0.3301,0.3967)

LN - Lognormal

Following the same procedures, we obtain for the other seasons

$$\begin{array}{lcl} X_A \sim G(\alpha_A, \beta_A) & \text{and} & Y_A \sim LN(\mu_A, \sigma_A), \\ X_{Sp} \sim B(k_{Sp}, c_{Sp}, \lambda_{Sp}) & \text{and} & Y_{Sp} \sim G(\alpha_{Sp}, \beta_{Sp}), \\ X_{Su} \sim G(\alpha_{Su}, \beta_{Su}) & \text{and} & Y_{Su} \sim W(\omega_{Su}, \delta_{Su}), \end{array}$$

where X_A and Y_A are, respectively, the observed and simulated wind speeds in Autumn, X_{Sp} and Y_{Sp}, respectively, the observed and simulated speed winds in Spring and X_{Su} and Y_{Su}, respectively, the observed and simulated speed winds in Summer for Castelo Branco's station.

Tab. 4.8: Fitted distributions to the observed, X, and simulated, Y, winds of Castelo Branco's station in the remaining seasons. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

		$\hat{\theta}_{MLE}$	CI (95%)			$\hat{\theta}_{MLE}$	CI (95%)
X_A				Y_A			
G	$\hat{\alpha}_A$	7.3327	(6.2194,8.8292)	LN	$\hat{\mu}_A$	1.7573	(1.7102,1.8061)
	$\hat{\beta}_A$	1.3447	(1.1327,1.6281)		$\hat{\sigma}_A$	0.3736	(0.3391,0.4063)
X_{Sp}				Y_{Sp}			
B	$\hat{k}_{Sp}$	1.3600	(0.8397,2.7629)	G	$\hat{\alpha}_{Sp}$	19.1999	(16.2291,23.0993)
	$\hat{c}_{Sp}$	5.2025	(4.3440,6.3834)		$\hat{\beta}_{Sp}$	2.4445	(2.0599,2.9435)
	$\hat{\lambda}_{Sp}$	0.1615	(0.1300,0.1847)		–	–	–
X_{Su}				Y_{Su}			
G	$\hat{\alpha}_{Su}$	16.5679	(14.0850,20.1388)	W	$\hat{\omega}_{Su}$	5.7901	(5.2909,6.4517)
	$\hat{\beta}_{Su}$	2.7476	(2.3260,3.3436)		$\hat{\delta}_{Su}$	8.6891	(8.4877,8.8877)

LN - Lognormal; G - Gamma; W - Weibull; B - Burr

In Fig. 4.6, we have the fitted marginal distributions plotted against the histograms of the data. For instance, if we look at the Spring, we can see the problem we mentioned above, that is, the simulated wind appears to be shifted to the right, when compared to the observed wind.

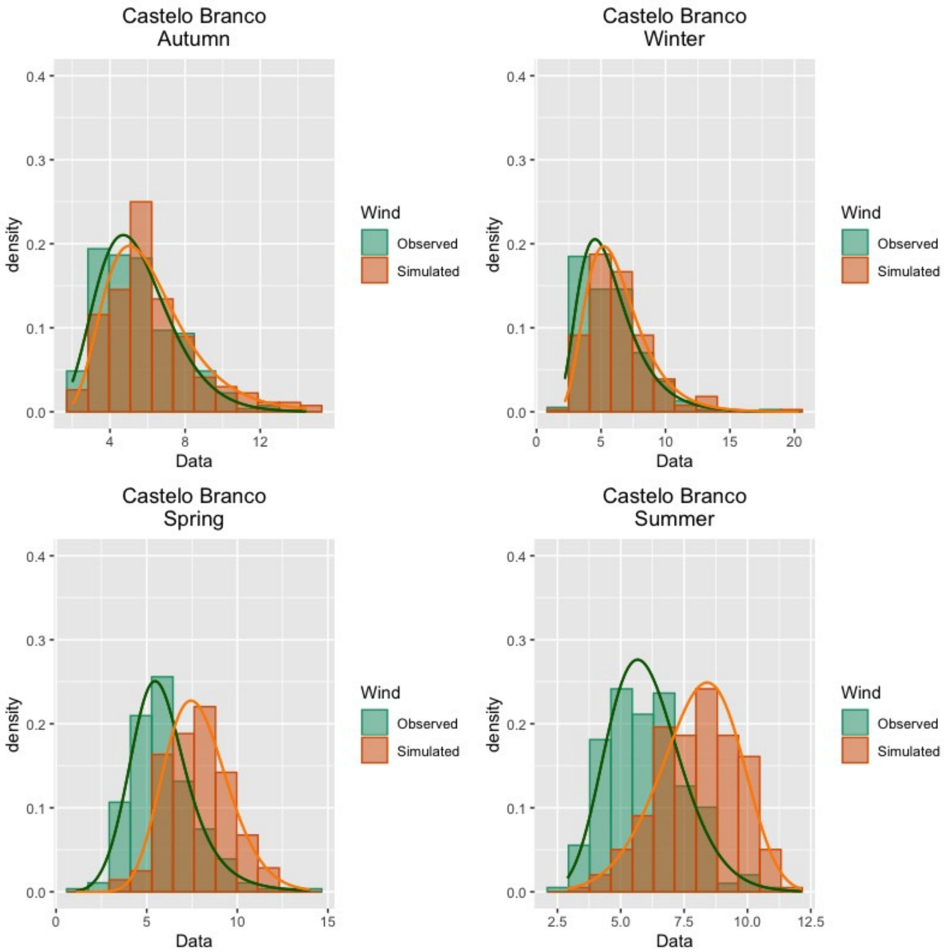


Fig. 4.6: Marginal distributions fitted to the 4 seasons of Castelo Branco.

Note that we just fitted the Weibull distribution to the simulated wind of the Summer. Moreover, the Weibull just proved to be the best distribution to fit the data 34 times out of 320 (40 stations $\times$ 4 seasons $\times$ 2 variables), which is about 11% of all the cases. Therefore, we acknowledge what [Çelik and Yilmaz, 2008] and [Mert and Karakus, 2015] state and, instead of simply fitting the Weibull without carrying out any analysis, we performed extensive goodness-of-fit tests considering 3 additional distributions. We can observe in Tab. 4.9 that the Lognormal was fitted the most in the Autumn and the Winter seasons while the 3-parameter Burr distribution was fitted the most in the Spring and the Summer.

Tab. 4.9: Percentage of the fitted distributions in the 40 stations per season.

Season	Lognormal			Gamma			Weibull			Burr		
	Overall	X	Y	Overall	X	Y	Overall	X	Y	Overall	X	Y
Autumn	**43.75%**	40%	47.5%	38.75%	40%	37.5%	1.25%	2.5%	0%	16.25%	17.5%	15%
Winter	**46.25%**	40%	52.5%	30%	32.5%	27.5%	6.25%	10%	2.5%	17.5%	17.5%	17.5%
Spring	17.5%	22.5%	12.5%	30%	27.5%	32.5%	15%	15%	15%	**37.5%**	10%	40%
Summer	13.75%	15%	12.5%	26.25%	20%	32.5%	20%	12.5%	27.5%	**40%**	52.5%	27.5%

The goodness-of-fit tests' results for the remaining seasons of Castelo Branco and for the other 8 selected meteorological stations can be found in Tab 4.10 (Autumn), Tab 4.11 (Winter), Tab 4.12 (Spring) and Tab 4.13 (Summer). Due to space limitations, "rej." refers to the rejection of H_0 and "not rej." to the non rejection of H_0, at a significance level of 5%. Tab 4.14 and Tab. 4.15 summarise the fitted distributions for the remaining 8 stations, for the observed and simulated wind data, respectively.

Tab. 4.10: Goodness-of-fit tests' results for the marginal distributions fitted to the selected 9 stations: Aveiro, Bragança, Castelo Branco, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo, in Autumn. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Autumn															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Aveiro	LN	*0.2256*	*not rej.*	–	–	*not rej.*	***987.9369***	*0.9940*	LN	*0.0352*	*not rej.*	–	–	*rej.*	***1068.3854***	*0.9916*
	G	0.1774	rej.	rej.	rej.	–	996.4002	0.9793	G	0.0051	rej.	rej.	rej.	–	1075.6898	0.9892
	W	0.0017	rej.	rej.	rej.	–	1028.0571	0.9263	W	0	rej.	rej.	rej.	–	1102.9757	0.9208
	B	0.1353	–	–	–	–	998.4580	0.9901	B	0.0024	–	–	–	–	1082.4329	0.9613
Bragança	LN	*0.0842*	*not rej.*	–	–	*not rej.*	***1048.1590***	*0.9947*	LN	0.2177	not rej.	–	–	rej.	1078.4558	0.9905
	G	0.0031	rej.	rej.	rej.	–	1056.7299	0.9921	G	*0.4431*	*not rej.*	*not rej.*	*not rej.*	–	***1076.5370***	*0.9960*
	W	0	rej.	rej.	rej.	–	1080.9091	0.9340	W	0.1813	not rej.	rej.	not rej.	–	1089.3010	0.9579
	B	0.0126	–	–	–	–	1060.5780	0.9781	B	0.2081	–	–	–	–	1085.3025	0.9952
Castelo Branco	LN	0.0651	not rej.	–	–	not rej.	**981.7391**	0.9951	LN	*0.5312*	*not rej.*	–	–	*not rej.*	***1042.8621***	*0.9954*
	G	*0.0166*	*not rej.*	*rej.*	*not rej.*	–	*986.0895*	*0.9962*	G	0.2117	not rej.	rej.	rej.	–	1048.6481	0.9758
	W	0	rej.	rej.	rej.	–	1010.0891	0.9380	W	0.0003	rej.	rej.	rej.	–	1080.0496	0.9450
	B	0.0093	–	–	–	–	995.1127	0.9845	B	0.6671	–	–	–	–	1047.8907	0.9882
Coruche	LN	0.1343	not rej.	–	–	not rej.	946.6190	0.9903	LN	0.3544	not rej.	–	–	not rej.	1018.0879	0.9939
	G	*0.3620*	*not rej.*	*not rej.*	*not rej.*	–	***943.2718***	*0.9941*	G	*0.5439*	*not rej.*	*not rej.*	*not rej.*	–	***1015.3763***	*0.9940*
	W	0.1018	not rej.	rej.	rej.	–	953.6941	0.9680	W	0.1403	not rej.	rej.	rej.	–	1029.1629	0.9722
	B	0.3163	–	–	–	–	950.3399	0.9904	B	0.4260	–	–	–	–	1021.4297	0.9943
Estremoz	LN	*0.2958*	*not rej.*	–	–	*not rej.*	***900.1120***	*0.9941*	LN	*0.0966*	*not rej.*	–	–	*not rej.*	***1014.1313***	*0.9927*
	G	0.1340	rej.	rej.	rej.	–	906.3282	0.9935	G	0.0123	rej.	rej.	rej.	–	1023.2105	0.9877
	W	0.0009	rej.	rej.	rej.	–	933.9832	0.9276	W	0	rej.	rej.	rej.	–	1053.9983	0.9195
	B	0.1696	–	–	–	–	913.2288	0.9735	B	0.0246	–	–	–	–	1024.7245	0.9698
Lisboa S1	LN	0.1548	not rej.	–	–	not rej.	932.3967	0.9889	LN	*0.1263*	*not rej.*	–	–	*not rej.*	***951.7668***	*0.9973*
	G	0.0441	rej.	rej.	rej.	–	935.8756	0.9730	G	0.0851	not rej.	not rej.	rej.	–	955.5905	0.9890
	W	0	rej.	rej.	rej.	–	967.6256	0.9492	W	0.0003	rej.	rej.	rej.	–	983.3038	0.9478
	B	*0.2537*	–	–	–	–	***931.7000***	*0.9919*	B	0.0755	–	–	–	–	958.6355	0.9890
Monção	LN	0.0409	rej.	–	–	rej.	958.8510	0.9409	LN	*0.2184*	*not rej.*	–	–	*not rej.*	***1041.9534***	*0.9946*
	G	0.3740	rej.	rej.	rej.	–	937.1490	0.9879	G	0.0943	not rej.	rej.	rej.	–	1050.7947	0.9617
	W	0.3269	not rej.	rej.	rej.	–	937.8651	0.9850	W	0.0001	rej.	rej.	rej.	–	1091.2221	0.9205
	B	*0.5506*	–	–	–	–	***931.4689***	*0.9892*	B	0.1012	–	–	–	–	1050.4775	0.9948
Sines	LN	*0.4551*	*not rej.*	–	–	*not rej.*	***1115.1722***	*0.9963*	LN	*0.4287*	*not rej.*	–	–	*rej.*	***1103.8004***	*0.9919*
	G	0.1985	rej.	rej.	rej.	–	1122.5696	0.9840	G	0.0510	not rej.	rej.	rej.	–	1111.6728	0.9795
	W	0.0005	rej.	rej.	rej.	–	1153.7464	0.9341	W	0	rej.	rej.	rej.	–	1144.1456	0.9191
	B	0.1610	–	–	–	–	1125.3915	0.9839	B	0.5196	–	–	–	–	1114.0867	0.9637
Vila do Bispo	LN	0.0229	rej.	–	–	rej.	1089.7207	0.9881	LN	*0.2880*	*not rej.*	–	–	*not rej.*	*1154.3784*	*0.9952*
	G	*0.0802*	*rej.*	*rej.*	*rej.*	–	***1086.9694***	*0.9910*	G	0.2275	not rej.	not rej.	not rej.	–	**1154.2928**	0.9917
	W	0.0191	not rej.	rej.	rej.	–	1099.3806	0.9596	W	0.0029	rej.	rej.	rej.	–	1176.0919	0.9624
	B	0.0436	-	-	-	-	1095.7837	0.9900	B	0.1478	–	–	–	–	1160.1564	0.9876

Tab. 4.11: Goodness-of-fit tests' results for the marginal distributions fitted to the selected 9 stations: Aveiro, Bragança, Castelo Branco, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo, in Winter. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Winter															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Aveiro	LN	*0.2785*	*not rej.*	–	–	*not rej.*	***1091.9658***	*0.9940*	LN	*0.0218*	*not rej.*	–	–	*rej.*	***1134.7641***	*0.9862*
	G	0.0252	rej.	rej.	rej.	–	1104.4129	0.9853	G	0.0113	rej.	rej.	rej.	–	1136.3760	0.9896
	W	0	rej.	rej.	rej.	–	1132.9759	0.9229	W	0.0004	rej.	rej.	rej.	–	1151.1594	0.9358
	B	0.1408	–	–	–	–	1102.1537	0.9787	B	0.0007	–	–	–	–	1149.0223	0.9858
Bragança	LN	0.7548	not rej.	–	–	not rej.	**1047.7036**	0.9980	LN	0.2304	rej.	–	–	rej.	**1079.9467**	0.9882
	G	*0.6085*	*not rej.*	*not rej.*	*not rej.*	–	*1052.2058*	*0.9911*	G	*0.2692*	*not rej.*	*rej.*	*rej.*	–	*1090.2489*	*0.9965*
	W	0.0396	rej.	rej.	rej.	–	1073.7886	0.9533	W	0.0600	not rej.	rej.	rej.	–	1093.7826	0.9455
	B	0.7010	–	–	–	–	1056.2373	0.9893	B	0.0843	–	–	–	–	1089.9912	0.9954
Castelo Branco	LN	*0.1278*	*not rej.*	–	–	*not rej.*	***1011.0699***	*0.9909*	LN	*0.6316*	*not rej.*	–	–	*not rej.*	***1024.9423***	*0.9960*
	G	0.0502	rej.	rej.	rej.	–	1017.6566	0.9682	G	0.4655	not rej.	not rej.	not rej.	–	1033.4992	0.9255
	W	0.0004	rej.	rej.	rej.	–	1045.6639	0.9261	W	0.0008	rej.	rej.	rej.	–	1073.8739	0.9394
	B	0.0174	–	–	–	–	1024.5307	0.9905	B	0.4363	–	–	–	–	1029.3653	0.9945
Coruche	LN	0.0535	rej.	–	–	rej.	977.1624	0.9361	LN	*0.3058*	*not rej.*	–	–	*not rej.*	***1008.8460***	*0.9971*
	G	0.5090	not rej.	not rej.	not rej.	–	956.3062	0.9843	G	0.2596	not rej.	not rej.	not rej.	–	1012.1822	0.9864
	W	0.7656	not rej.	rej.	not rej.	–	959.3747	0.9600	W	0.0037	rej.	rej.	rej.	–	1038.4857	0.9495
	B	*0.5860*	–	–	–	–	***955.7600***	*0.9859*	B	0.2646	–	–	–	–	1015.9681	0.9970
Estremoz	LN	0.3725	not rej.	–	–	not rej.	**848.2808**	0.9917	LN	*0.0218*	*not rej.*	–	–	*not rej.*	***919.6256***	*0.9925*
	G	*0.3468*	*not rej.*	*not rej.*	*not rej.*	–	*853.9638*	*0.9795*	G	0.0132	not rej.	rej.	rej.	–	926.5752	0.9671
	W	0.0268	rej.	rej.	rej.	–	879.1227	0.9266	W	0.0001	rej.	rej.	rej.	–	956.0387	0.9251
	B	0.1736	–	–	–	–	860.7098	0.9901	B	0.0017	–	–	–	–	931.8784	0.9873
Lisboa S1	LN	0.0111	not rej.	–	–	rej.	1022.6113	0.9311	LN	*0.2168*	*not rej.*	–	–	*not rej.*	***1006.7846***	*0.9933*
	G	0.0114	not rej.	rej.	rej.	–	1008.4108	0.9857	G	0.1209	not rej.	not rej.	not rej.	–	1007.9534	0.9883
	W	0	rej.	rej.	rej.	–	1027.1327	0.9284	W	0.0011	rej.	rej.	rej.	–	1029.6724	0.9516
	B	*0.0302*	–	–	–	–	***1004.4609***	*0.9917*	B	0.1951	–	–	–	–	1015.1921	0.9791
Monção	LN	0.0010	rej.	–	–	rej.	943.1875	0.8522	LN	*0.5797*	*not rej.*	–	–	*not rej.*	***977.4896***	*0.9939*
	G	0.1256	not rej.	rej.	rej.	–	901.7333	0.9708	G	0.1899	not rej.	rej.	rej.	–	984.593	0.9855
	W	0.1005	not rej.	rej.	rej.	–	896.9995	0.9238	W	0.0001	rej.	rej.	rej.	–	1017.6106	0.9209
	B	*0.3835*	–	–	–	–	***888.2799***	*0.9783*	B	0.5595	–	–	–	–	985.5430	0.9788
Sines	LN	*0.0122*	*not rej.*	–	–	*rej.*	***1072.7596***	*0.9882*	LN	*0.6708*	*not rej.*	–	–	*not rej.*	***1098.4445***	*0.9965*
	G	0.0003	rej.	rej.	rej.	–	1082.5850	0.9827	G	0.3566	not rej.	rej.	not rej.	–	1105.2663	0.9871
	W	0	rej.	rej.	rej.	–	1114.8204	0.9072	W	0.0012	rej.	rej.	rej.	–	1139.1605	0.9286
	B	0.0234	–	–	–	–	1084.4619	0.9613	B	0.6198	–	–	–	–	1108.1407	0.9861
Vila do Bispo	LN	0.1912	not rej.	–	–	rej.	1017.9718	0.9869	LN	*0.1593*	*not rej.*	–	–	*not rej.*	***1089.8688***	*0.9957*
	G	0.5840	not rej.	not rej.	not rej.	–	**1014.1752**	0.9902	G	0.0695	not rej.	not rej.	rej.	–	1091.8376	0.9947
	W	*0.6607*	*not rej.*	*rej.*	*not rej.*	–	*1023.4748*	*0.9630*	W	0.0003	rej.	rej.	rej.	–	1113.7424	0.9528
	B	0.5398	-	-	-	-	1021.3486	0.9915	B	0.0777	–	–	–	–	1098.2730	0.9807

Tab. 4.12: Goodness-of-fit tests' results for the marginal distributions fitted to the selected 9 stations: Aveiro, Bragança, Castelo Branco, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo, in Spring. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Spring															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Aveiro	LN	*0.2785*	*not rej.*	–	–	*not rej.*	***1091.9658***	*0.9940*	LN	*0.0218*	*not rej.*	–	–	*rej.*	***1134.7641***	*0.9862*
	G	0.0578	not rej.	rej.	rej.	–	949.4142	0.9733	G	0.0610	rej.	rej.	rej.	–	1038.3216	0.9854
	W	0	rej.	rej.	rej.	–	985.9209	0.9392	W	0	rej.	rej.	rej.	–	1075.5987	0.9270
	B	0.2205	–	–	–	–	**945.5850**	0.9951	B	0.0336	–	–	–	–	1041.2832	0.9893
Bragança	LN	*0.9360*	*not rej.*	–	–	*not rej.*	***881.5242***	*0.9960*	LN	0.0522	not rej.	–	–	rej.	1009.0563	0.9744
	G	0.7582	not rej.	not rej.	not rej.	–	883.2449	0.9938	G	0.2101	not rej.	not rej.	not rej.	–	998.4797	0.9864
	W	0.0030	rej.	rej.	rej.	–	913.0884	0.9493	W	0.0717	not rej.	rej.	rej.	–	1000.5703	0.9905
	B	0.6907	–	–	–	–	889.0241	0.9886	B	*0.3222*	–	–	–	–	***997.7914***	*0.9866*
Castelo Branco	LN	0.2967	not rej.	–	–	not rej.	965.0268	0.9941	LN	0.0001	rej.	–	–	rej.	963.7642	0.9529
	G	0.3770	not rej.	not rej.	not rej.	–	959.3523	0.9741	G	*0.0018*	*not rej.*	*rej.*	*rej.*	–	***956.7894***	*0.9884*
	W	0.0326	not rej.	rej.	rej.	–	985.8117	0.9550	W	0.0888	not rej.	not rej.	not rej.	–	960.9676	0.9940
	B	*0.1248*	–	–	–	–	***957.5969***	*0.9965*	B	0.0267	–	–	–	–	959.2173	0.9940
Coruche	LN	0.0241	rej.	–	–	rej.	894.6108	0.9475	LN	0.0024	rej.	–	–	rej.	895.9789	0.9635
	G	0.1334	rej.	rej.	rej.	–	879.3434	0.9941	G	0.0171	rej.	rej.	rej.	–	887.1257	0.9787
	W	0.0588	rej.	rej.	rej.	–	875.8260	0.9867	W	0.1230	not rej.	rej.	rej.	–	886.6079	0.9911
	B	*0.4291*	–	–	–	–	***868.8189***	*0.9928*	B	*0.2379*	–	–	–	–	***876.8867***	*0.9800*
Estremoz	LN	0.1413	not rej.	–	–	not rej.	**826.0191**	0.9968	LN	0.1378	rej.	–	–	rej.	931.8996	0.9841
	G	*0.1854*	*not rej.*	*not rej.*	*not rej.*	–	*826.7676*	*0.9935*	G	0.3994	not rej.	not rej.	not rej.	–	**926.0302**	0.9803
	W	0.0041	not rej.	rej.	rej.	–	853.0373	0.9523	W	*0.5057*	*not rej.*	*rej.*	*rej.*	–	*928.5864*	*0.9762*
	B	0.0789	–	–	–	–	834.8899	0.9938	B	0.5754	–	–	–	–	928.8604	0.9899
Lisboa S1	LN	*0.7685*	*not rej.*	–	–	*not rej.*	***839.5842***	*0.9966*	LN	0.3092	not rej.	–	–	not rej.	948.4956	0.9907
	G	0.6523	not rej.	not rej.	not rej.	–	842.1768	0.9743	G	*0.5487*	*not rej.*	*not rej.*	*not rej.*	–	***944.7238***	*0.9954*
	W	0.0020	rej.	rej.	rej.	–	881.3416	0.9417	W	0.1377	not rej.	rej.	rej.	–	963.4104	0.9760
	B	0.4807	–	–	–	–	847.6993	0.9924	B	0.6435	–	–	–	–	947.4275	0.9977
Monção	LN	*0.0551*	*rej.*	–	–	*rej.*	*899.1556*	*0.8663*	LN	*0.0172*	*rej.*	–	–	*not rej.*	***942.1138***	*0.9917*
	G	0.0702	rej.	rej.	rej.	–	871.2565	0.9912	G	0.0026	rej.	rej.	rej.	–	946.6762	0.9919
	W	0	rej.	rej.	rej.	–	877.4324	0.8825	W	0	rej.	rej.	rej.	–	985.1072	0.9228
	B	0.0703	–	–	–	–	**862.3487**	0.9801	B	0.0307	–	–	–	–	949.4112	0.9746
Sines	LN	0.1203	not rej.	–	–	rej.	957.7869	0.9900	LN	0.1052	not rej.	–	–	not rej.	**927.0724**	0.9971
	G	*0.0841*	*not rej.*	*not rej.*	*not rej.*	–	***955.8202***	*0.9790*	G	*0.0796*	*not rej.*	*not rej.*	*not rej.*	–	*928.1274*	*0.9950*
	W	0.0015	not rej.	rej.	rej.	–	967.6011	0.9571	W	0.0003	rej.	rej.	rej.	–	954.5955	0.9464
	B	0.0088	–	–	–	–	966.1275	0.9797	B	0.0358	–	–	–	–	937.5305	0.9888
Vila do Bispo	LN	0.0001	rej.	–	–	rej.	1065.4566	0.9787	LN	0.0173	not rej.	–	–	rej.	1066.0528	0.9821
	G	0.0007	rej.	rej.	rej.	–	**1057.2428**	0.9831	G	0.1088	not rej.	not rej.	not rej.	–	1058.6560	0.9924
	W	*0.0016*	*not rej.*	*rej.*	*rej.*	–	*1057.8196*	*0.9795*	W	0.2569	not rej.	rej.	not rej.	–	1065.4495	0.9868
	B	0.0013	–	–	–	–	1058.7209	0.9960	B	*0.3395*	–	–	–	–	***1058.3938***	*0.9936*

Tab. 4.13: Goodness-of-fit tests' results for the marginal distributions fitted to the selected 9 stations: Aveiro, Bragança, Castelo Branco, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo, in Summer. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Summer															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Aveiro	LN	0.4107	not rej.	–	–	not rej.	**818.4938**	0.9915	LN	*0.0083*	*not rej.*	–	–	*rej.*	***1054.3404***	*0.9891*
	G	0.1932	rej.	rej.	rej.	–	824.0646	0.9864	G	0.0042	not rej.	rej.	rej.	–	1057.8199	0.9820
	W	0	rej.	rej.	rej.	–	865.0809	0.9156	W	0	rej.	rej.	rej.	–	1083.2003	0.9240
	B	*0.4689*	–	–	–	–	*826.0365*	*0.9791*	B	0.0005	–	–	–	–	1070.5211	0.9569
Bragança	LN	0.0167	not rej.	–	–	not rej.	**837.8349**	0.9930	LN	0.0002	rej.	–	–	rej.	926.2814	0.9575
	G	0.0527	not rej.	not rej.	not rej.	–	839.1689	0.9721	G	0.0097	rej.	rej.	rej.	–	912.8089	0.9885
	W	0.0007	rej.	rej.	rej.	–	873.6204	0.9506	W	*0.7108*	*not rej.*	*not rej.*	*not rej.*	–	***900.8988***	*0.9930*
	B	*0.0524*	–	–	–	–	*841.4688*	*0.9966*	B	0.582	–	–	–	–	901.0611	0.9939
Castelo Branco	LN	0.2300	not rej.	–	–	rej.	853.7916	0.9684	LN	0.0356	not rej.	–	–	rej.	923.6343	0.9811
	G	*0.2967*	*not rej.*	*not rej.*	*not rej.*	–	***853.0971***	*0.9938*	G	*0.0628*	*not rej.*	*not rej.*	*not rej.*	–	*910.9966*	*0.9667*
	W	0.0033	rej.	rej.	rej.	–	873.9541	0.9523	W	0.0071	not rej.	rej.	rej.	–	**893.8864**	0.9821
	B	0.3182	–	–	–	–	862.7156	0.9953	B	0.0153	–	–	–	–	897.9155	0.9892
Coruche	LN	0	rej.	–	–	rej.	727.8740	0.8806	LN	0.0234	rej.	–	–	rej.	819.8946	0.9746
	G	0.0008	rej.	rej.	rej.	–	708.4084	0.9496	G	0.0645	rej.	rej.	rej.	–	813.0878	0.9854
	W	0.0002	rej.	rej.	rej.	–	703.0534	0.9557	W	0.1426	not rej.	not rej.	not rej.	–	808.0615	0.9923
	B	*0.1332*	–	–	–	–	***677.2790***	*0.9603*	B	*0.1974*	–	–	–	–	***807.2367***	*0.9941*
Estremoz	LN	*0.5307*	*not rej.*	–	–	*not rej.*	***697.3182***	*0.9870*	LN	0	rej.	–	–	rej.	872.7954	0.9159
	G	0.4311	not rej.	not rej.	not rej.	–	702.0875	0.8492	G	0	rej.	rej.	rej.	–	855.7131	0.9156
	W	0	rej.	rej.	rej.	–	762.7177	0.9309	W	*0.2564*	*not rej.*	*rej.*	*not rej.*	–	***812.5021***	*0.9879*
	B	0.2119	–	–	–	–	701.5615	0.9776	B	0.1769	–	–	–	–	815.1272	0.9830
Lisboa S1	LN	0.8157	not rej.	–	–	not rej.	762.6215	0.9970	LN	0.7835	not rej.	–	–	not rej.	879.7004	0.9953
	G	*0.7854*	*not rej.*	*not rej.*	*not rej.*	–	***762.4808***	*0.9967*	G	*0.7361*	*not rej.*	*not rej.*	*not rej.*	–	***879.0254***	*0.9878*
	W	0.0294	rej.	rej.	rej.	–	784.8985	0.9541	W	0.0235	rej.	rej.	rej.	–	897.5718	0.9579
	B	0.5606	–	–	–	–	771.5239	0.9910	B	0.4683	–	–	–	–	889.0989	0.9793
Monção	LN	*0.1849*	*not rej.*	–	–	*rej.*	***765.5719***	*0.9921*	LN	*0.9073*	*not rej.*	–	–	*not rej.*	***820.8076***	*0.9942*
	G	0.0554	rej.	rej.	rej.	–	769.1744	0.9833	G	0.7479	not rej.	rej.	not rej.	–	824.8959	0.9859
	W	0	rej.	rej.	rej.	–	800.1449	0.9261	W	0.0008	rej.	rej.	rej.	–	867.4224	0.9222
	B	0.1726	–	–	–	–	777.0479	0.9624	B	0.9222	–	–	–	–	827.3061	0.9841
Sines	LN	0.0004	rej.	–	–	rej.	929.0852	0.9638	LN	0.0171	rej.	–	–	rej.	929.0852	0.9788
	G	0.0044	rej.	rej.	rej.	–	927.3577	0.9272	G	*0.0238*	*not rej.*	*rej.*	*rej.*	–	***927.3577***	*0.9558*
	W	*0.0641*	*rej.*	*rej.*	*rej.*	–	***935.4516***	*0.9778*	W	0.0019	rej.	rej.	rej.	–	935.4516	0.9413
	B	0.0388	–	–	–	–	938.2671	0.9633	B	0.0021	–	–	–	–	938.2671	0.9736
Vila do Bispo	LN	0	rej.	–	–	rej.	1136.5406	0.9472	LN	0.0124	rej.	–	–	rej.	1042.5134	0.9618
	G	0.0001	rej.	rej.	rej.	–	1121.5068	0.9221	G	0.1103	rej.	rej.	rej.	–	1030.5644	0.9827
	W	*0.0801*	*rej.*	*rej.*	*rej.*	–	***1100.3306***	*0.9797*	W	*0.8020*	*not rej.*	*not rej.*	*not rej.*	–	***1018.6280***	*0.9949*
	B	0.0450	–	–	–	–	1111.6229	0.9561	B	0.5305	–	–	–	–	1022.2693	0.9945

Tab. 4.14: Fitted distributions to the **observed wind** of the remaining 8 stations: Aveiro, Bragança, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

Autumn				Winter				Spring				Summer			
		$\hat{\theta}_{MLE}$	CI (95%)			$\hat{\theta}_{MLE}$	CI (95%)			$\hat{\theta}_{MLE}$	CI (95%)			$\hat{\theta}_{MLE}$	CI (95%)
LN	$\hat{\mu}_{Av}$	1.6216	(1.5726, 1.6703)	LN	$\hat{\mu}_{Av}$	1.7035	(1.6552, 1.7638)	LN	$\hat{\mu}_{Av}$	1.8313	(1.7954, 1.8699)	B	$\hat{k}_{Av}$	0.7421	(0.4887, 1.1888)
	$\hat{\sigma}_{Av}$	0.3914	(0.3548, 0.4261)		$\hat{\sigma}_{Av}$	0.4642	(0.4222, 0.5040)		$\hat{\sigma}_{Av}$	0.2985	(0.2710, 0.3246)		$\hat{c}_{Av}$	8.6840	(7.2000, 10.9152)
	–	–	–		–	–		–	–	–	–		$\hat{\lambda}_{Av}$	0.1769	(0.1612, 0.1901)
LN	$\hat{\mu}_B$	1.5692	(1.5102, 1.6285)	LN	$\hat{\mu}_B$	4.835	(4.1001, 5.8293)	LN	$\hat{\mu}_B$	1.7654	(1.7320, 1.7984)	B	$\hat{k}_B$	1.3231	(0.8288, 2.6632)
	$\hat{\sigma}_B$	0.4646	(0.4217, 0.5066)		$\hat{\sigma}_B$	0.8826	(0.7409, 1.0742)		$\hat{\sigma}_B$	0.2637	(0.2399, 0.2881)		$\hat{c}_B$	6.2980	(5.2466, 7.6926)
	–	–	–		–	–	–		–	–	–		$\hat{\lambda}_B$	0.1732	(0.1457, 0.1934)
G	$\hat{\alpha}_C$	5.1743	(4.4128, 6.3163)	B	$\hat{k}_C$	3.6282	(1.7408, 61.9258)	B	$\hat{k}_C$	2.7836	(1.4599, 15.5809)	B	$\hat{k}_C$	2.1764	(1.2565, 7.2899)
	$\hat{\beta}_C$	0.9889	(0.8344, 1.2206)		$\hat{c}_C$	2.9924	(2.5501, 3.5648)		$\hat{c}_C$	5.8422	(4.9500, 7.0183)		$\hat{c}_C$	8.5089	(7.1399, 10.1859)
	–	–	–		$\hat{\lambda}_C$	0.1220	(0.0378, 0.1709)		$\hat{\lambda}_C$	0.1274	(0.0852, 0.1487)		$\hat{\lambda}_C$	0.1498	(0.1227, 0.1644)
LN	$\hat{\mu}_E$	1.5288	(1.4852, 1.5761)	G	$\hat{\alpha}_E$	6.9393	(5.8195, 8.4359)	G	$\hat{\alpha}_E$	14.5757	(12.4803, 17.7280)	LN	$\hat{\mu}_E$	1.5900	(1.9067, 1.9615)
	$\hat{\sigma}_E$	0.3531	(0.3213, 0.3838)		$\hat{\beta}_E$	1.3952	(1.1709, 1.7132)		$\hat{\beta}_E$	2.7966	(2.3832, 3.4161)		$\hat{\sigma}_E$	0.2105	(0.1958, 0.2333)
B	$\hat{k}_L$	0.9793	(0.6204, 1.7559)	B	$\hat{k}_L$	1.2991	(0.7990, 2.6257)	LN	$\hat{\mu}_L$	1.8613	(1.8336, 1.8874)	G	$\hat{\alpha}_L$	26.2132	(22.0828, 32.1281)
	$\hat{c}_L$	5.3047	(4.3552, 6.6461)		$\hat{c}_L$	4.4700	(3.7404, 5.5092)		$\hat{\sigma}_L$	0.2114	(0.1921, 0.2299)		β_L	3.8139	(3.2083, 4.6797)
	$\hat{\lambda}_L$	0.1706	(0.1428, 0.1943)		$\hat{\lambda}_L$	0.1567	(0.1219, 0.1828)		–	–	–		–	–	–
B	$\hat{k}_M$	3.0982	(1.5767, 36.3077)	B	$\hat{k}_M$	3.0463	(1.5317, 30.5305)	LN	$\hat{\mu}_M$	1.6721	(1.6344, 1.7102)	LN	$\hat{\mu}_M$	1.6754	(1.6471, 1.7042)
	$\hat{c}_M$	3.5241	(2.9900, 4.2089)		$\hat{c}_M$	3.5296	(2.9807, 4.2666)		$\hat{\sigma}_M$	0.3029	(0.2762, 0.3297)		$\hat{\sigma}_M$	0.2322	(0.2105, 0.2526)
	$\hat{\lambda}_M$	0.1353	(0.0546, 0.1759)		$\hat{\lambda}_M$	0.1341	(0.0567, 0.1760)		–	–	–		–	–	–
LN	$\hat{\mu}_S$	1.9083	(1.8614, 1.9575)	LN	$\hat{\mu}_S$	1.9079	(1.8613, 1.9552)	G	$\hat{\alpha}_S$	13.8738	(11.6967, 16.8954)	W	$\hat{\omega}_S$	4.8835	(4.4309, 5.4003)
	$\hat{\sigma}_S$	0.3742	(0.3394, 0.4068)		$\hat{\sigma}_S$	0.3558	(0.3222, 0.3870)		$\hat{\beta}_S$	1.8287	(1.5325, 2.2304)		$\hat{\delta}_S$	8.5268	(8.2824, 8.7600)
G	$\hat{\alpha}_{VB}$	7.6060	(6.4381, 9.1430)	W	$\hat{\omega}_{VB}$	3.0407	(2.7539, 3.3825)	W	$\hat{\omega}_{VB}$	4.0879	(3.7286, 4.5371)	W	$\hat{\omega}_{VB}$	4.1212	(3.7543, 4.5828)
	$\hat{\beta}_{VB}$	1.0877	(0.9138, 1.3168)		$\hat{\delta}_{VB}$	8.1365	(7.7695, 8.519)		$\hat{\delta}_{VB}$	8.7183	(8.4324, 8.9975)		$\hat{\delta}_{VB}$	9.5870	(9.2733, 9.8991)

Tab. 4.15: Fitted distributions to the **simulated wind** of the remaining 8 stations: Aveiro, Bragança, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

Autumn				**Winter**				**Spring**				**Summer**			
		$\hat{\theta}_{MLE}$	CI (95%)			$\hat{\theta}_{MLE}$	CI (95%)			$\hat{\theta}_{MLE}$	CI (95%)			$\hat{\theta}_{MLE}$	CI (95%)
LN	$\hat{\mu}_{Av}$	1.8695	(1.8208,1.9163)	LN	$\hat{\mu}_{Av}$	1.9382	(1.8877,1.9898)	LN	$\hat{\mu}_{Av}$	2.1014	(2.0658,2.1359)	LN	$\hat{\mu}_{Av}$	2.1101	(2.0751,2.1445)
	$\hat{\sigma}_{Av}$	0.3627	(0.3303,0.3944)		$\hat{\sigma}_{Av}$	0.4027	(0.3654,0.4380)		$\hat{\sigma}_{Av}$	0.2741	(0.2485,0.2985)		$\hat{\sigma}_{Av}$	0.2715	(0.2465,0.2950)
G	$\hat{\alpha}_B$	5.6890	(4.8003,6.8790)	G	$\hat{\alpha}_B$	5.0187	(4.2599,6.0802)	B	$\hat{k}_B$	3.0920	(1.5931,29.2842)	W	$\hat{\omega}_B$	4.7888	(4.3512,5.3332)
	$\hat{\beta}_B$	0.9464	(0.7946,1.1577)		$\hat{\beta}_B$	0.8489	(0.7123,1.0384)		$\hat{c}_B$	4.8306	(4.1060,5.7766)		$\hat{\delta}_B$	7.3762	(7.1682,7.5721)
	–	–	–		–	–	–		$\hat{\lambda}_B$	0.1044	(0.0569,0.1259)		–	–	–
G	$\hat{\alpha}_C$	6.3357	(5.3450,7.7181)	LN	$\hat{\mu}_C$	1.8516	(1.8043,1.9006)	B	$\hat{k}_C$	3.0240	(1.5641,26.9446)	B	$\hat{k}_C$	4.5180	(2.0329,105.4045)
	$\hat{\beta}_C$	0.9373	(0.7872,1.1433)		$\hat{\sigma}_C$	0.3652	(0.3306,0.3981)		$\hat{c}_C$	7.9472	(6.7367,9.5330)		$\hat{c}_C$	8.2577	(7.0868,9.8016)
	–	–	–		–	–	–		$\hat{\lambda}_C$	0.0956	(0.0665,0.1073)		$\hat{\lambda}_C$	0.0915	(0.0576,0.1040)
LN	$\hat{\mu}_E$	1.7152	(1.6636,1.7633)	LN	$\hat{\mu}_E$	1.7452	(1.6978,1.7973)	W	$\hat{\omega}_E$	4.8751	(4.4485,5.4057)	W	$\hat{\omega}_E$	6.3563	(5.7932,7.0696)
	$\hat{\sigma}_E$	0.3735	(0.3400,0.4065)		$\hat{\sigma}_E$	0.3662	(0.3308,0.4005)		$\hat{\delta}_E$	7.8695	(7.6574,8.0838)		$\hat{\delta}_E$	7.6126	(7.4357,7.7705)
LN	$\hat{\mu}_L$	1.8269	(1.7812,1.8719)	LN	$\hat{\mu}_L$	1.8654	(1.8245,1.9108)	G	$\hat{\alpha}_L$	21.6033	(18.3379,26.0382)	G	$\hat{\alpha}_L$	22.2501	(18.7282,27.0391)
	$\hat{\sigma}_L$	0.3389	(0.3047,0.3689)		$\hat{\sigma}_L$	0.3346	(0.3036,0.3655)		$\hat{\beta}_L$	2.7074	(2.2913,3.2685)		$\hat{\beta}_L$	2.6995	(2.2713,3.2882)
LN	$\hat{\mu}_M$	2.0059	(1.9649,2.0460)	LN	$\hat{\mu}_M$	2.0374	(1.9949,2.0777)	LN	$\hat{\mu}_M$	2.0697	(2.0411,2.0976)	LN	$\hat{\mu}_M$	2.0300	(2.0065,2.0535)
	$\hat{\sigma}_M$	0.3203	(0.2903,0.3495)		$\hat{\sigma}_M$	0.3067	(0.2769,0.3353)		$\hat{\sigma}_M$	0.2229	(0.2030,0.2420)		$\hat{\sigma}_M$	0.1834	(0.1666,0.2001)
LN	$\hat{\mu}_S$	2.0100	(1.9686,2.0532)	LN	$\hat{\mu}_S$	2.0591	(2.0181,2.1009)	G	$\hat{\alpha}_S$	22.0165	(18.6476,26.7872)	G	$\hat{\alpha}_S$	26.0509	(21.8801,31.4503)
	$\hat{\sigma}_S$	0.3300	(0.3007,0.3590)		$\hat{\sigma}_S$	0.3232	(0.2926,0.3509)		$\hat{\beta}_S$	2.4718	(2.0901,3.0253)		$\hat{\beta}_S$	3.0419	(2.5512,3.6847)
LN	$\hat{\mu}_{VB}$	2.0774	(2.0322,2.1215)	LN	$\hat{\mu}_{VB}$	2.1350	(2.0889,2.1809)	B	$\hat{k}_{VB}$	2.9340	(1.5485,23.4852)	W	$\hat{\omega}_{VB}$	5.2408	(4.7736,5.8157)
	$\hat{\sigma}_{VB}$	0.3504	(0.3180,0.3816)		$\hat{\sigma}_{VB}$	0.3493	(0.3148,0.3819)		$\hat{c}_{VB}$	5.7342	(4.8538,6.8178)		$\hat{\delta}_{VB}$	10.2114	(9.9545,10.4663)
	–	–	–		–	–	–		$\hat{\lambda}_{VB}$	0.0869	(0.0530,0.1016)		–	–	–

4.4.2 Copulas

The copulas fitted to the observed and to the simulated winds will be the ones mentioned in Chapter 2. Occasionally, the Joe copula, an Archimedean copula, will be fitted. Its expression and characteristics can be found in Appendix A. The *VineCopula* package in R provides a useful function, *BiCopSelect*, which allows to select the best copula among 40 family of copulas according to 3 criteria: the AIC, the BIC or the log-likelihood. The function includes the following copulas and their codes:

- 0: Product, Π
- 1: Gaussian, C_ρ^{gaus}
- 2: Student t, $C_{\rho\eta}^t$
- 3: Clayton, C_α^c
- 4: Gumbel, C_α^{gu}
- 5: Frank, C_α^f
- 6: Joe, C_α^j
- 7: BB1
- 8: BB6
- 9: BB7
- 10: BB8
- 13: Survival Clayton, C_α^{sc}
- 14: Survival Gumbel, C_α^{sg}
- 16: Survival Joe
- 17: Survival BB1
- 18: Survival BB6
- 19: Survival BB7
- 20: Survival BB8
- 23: Rot. 90º Clayton
- 24: Rot. 90º Gumbel
- 26: Rot. 90º Joe
- 27: Rot. 90º BB1
- 28: Rot. 90º BB6
- 29: Rot. 90º BB7
- 30: Rot. 90º BB8
- 33: Rot. 270º Clayton
- 34: Rot. 270º Gumbel
- 36: Rot. 270º Joe
- 37: Rot. 270º BB1
- 38: Rot. 270º BB6
- 39: Rot. 270º BB7
- 40: Rot. 270º BB8
- 104: Tawn Type 1
- 114: Rotated 180º Tawn Type 1
- 124: Rotated 90º Tawn Type 1
- 134: Rotated 270º Tawn Type 1
- 204: Tawn Type 2
- 214: Rotated 180º Tawn Type 2
- 224: Rotated 90º Tawn Type 2
- 234: Rotated 270º Tawn Type 2

In our study, the copulas selected were Gaussian (1), Student t (2), Clayton (3), Gumbel (4), Frank (5), Joe (6), BB1 (7), BB7 (9), BB8 (10), Survival Clayton (13), Survival Gumbel (14), Survival BB1 (17), Survival BB7 (19), Survival BB8 (20), Tawn Type 1 (104) and Tawn Type 2 (204). However, through the function *mvdc* of the *copula* package, we were not able to simulate a bivariate distribution from the copulas BB1, BB7, BB8, Survival BB1, Survival BB7, Survival BB8, Tawn Type 1 and Tawn Type 2, and we were not able to find an alternative function. Therefore, as simulating a bivariate distribution from the copula was one of our objectives, we decided to select the copula with the lowest AIC among the first 6 families of copulas plus the corresponding survival copulas, that is from the families 0 to 6 and from 13 to 16. Globally, we did not find a major difference when fitting the "alternative" copulas, since they capture the same tail dependence and they were not rejected, at a level of significance of 5%, by any of the goodness-of-fit tests performed.

The *BiCopSelect* function also estimates the copula parameters by maximum pseudo-likelihood, if u and v are the pseudo-observations of a given bivariate vector (X,Y), by the IFM method, if $u = F(X)$ and $v = G(Y)$, or by the estimation based on Kendall's tau. We chose to estimate the copula parameters by MPLE. Later, we will compare the different types of estimation we mentioned in Chapter 3. To assess the fit of the copula to the data, we will base ourselves on the conclusions of [Genest *et al*, 2009] stated in Section 3.2 and perform 3 goodness-of-fit tests: the Cramér-von-Mises test statistic based on the Rosenblatt's transformation, $S_n^{(B)}$, the rank-based version of Cramér-von-Mises using the empirical copula, S_n, and the Cramér-von-Mises test statistic based on Kendall's process, $S_n^{(K)}$. We will continue to present the complete procedure for the Winter observed and simulated data sets of Castelo Branco and the overall results for the remaining seasons and stations.

Through the AIC, we came to the conclusion that the copula which best fits Castelo Branco's Winter data is the Gumbel. As stated in Section 2.5, this copula belongs to the Archimedean family and it is characterised for having upper tail dependence. Therefore, it is most used when the dependence between high values of the two univariate distributions is stronger than between their low values. In our case, we conclude that higher values of the observed wind are associated with higher values of the simulated wind and thus the simulated data seems to match strong winds really well, relatively well moderate winds and not so well weak winds.

The estimated copula parameter by maximum pseudo-likelihood and its confidence interval, the tail dependence coefficients and the p-values of the goodness-of-fit copulas tests, $S_n^{(B)}$, S_n and $S_n^{(K)}$ given in (3.39), (3.32) and (3.35), respectively, are presented in Tab. 4.16.

Tab. 4.16: Copula selected according to the AIC criterion to fit the Winter season data of Castelo Branco and copula parameter estimate ($\hat{\alpha}$) obtained through the MPLE method, the estimated coefficients of the tail dependence and the p-values of the goodness-of-fit copula tests performed, $S_n^{(B)}$, S_n and $S_n^{(K)}$.

Copula	$\hat{\theta}_{MPLE}$		Tail Dependence		p-values		
	$\hat{\alpha}$	CI (95%)	λ_L	$\hat{\lambda}_U$	$S_n^{(B)}$	S_n	$S_n^{(K)}$
C_α^{gu}	2.2604	(1.9783, 2.5424)	0	0.6411	0.965	1	0.530

The null hyphotesis, $H_0 : C \in C_0$ where C_0 is the Gumbel Copula, is clearly not rejected by any of the tests statistics. Thus, the Gumbel Copula seems to fit well the dependence between the observed and the simulated wind speed values of Castelo Branco during the Winter season. The probability density function of the fitted Gumbel Copula can be found in the upper right corner of Fig. 4.7.

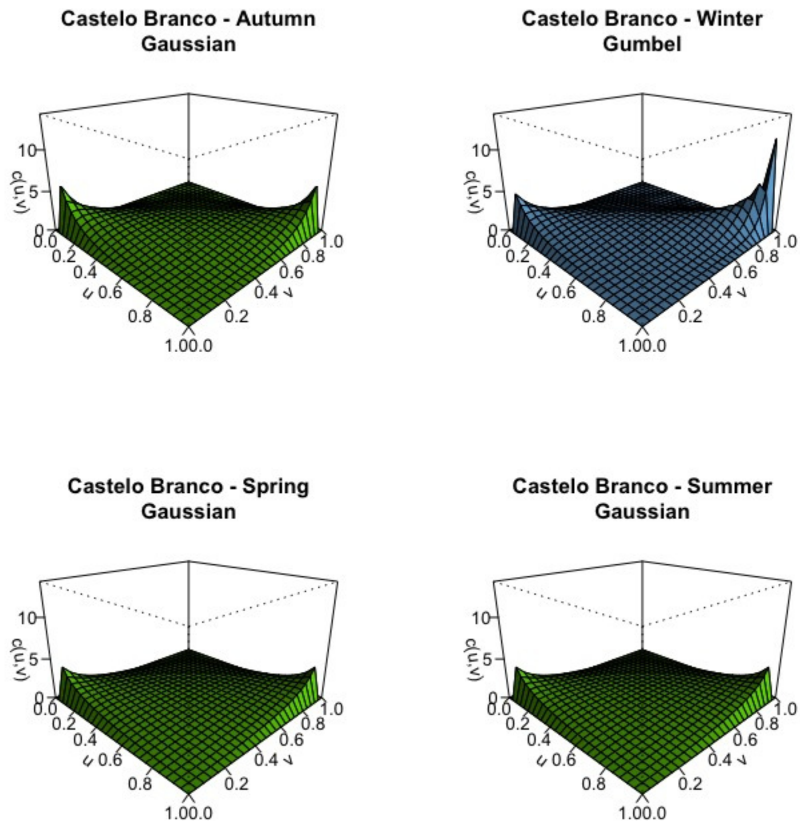


Fig. 4.7: Probability density functions of the copulas fitted to Castelo Branco's data.

The copula, its parameters estimates, its tail dependence coefficients and the p-values of the 3 goodness-of-fit tests for the remaining seasons of Castelo Branco are presented on Tab. 4.17. For all three seasons, the Gaussian Copula was found to fit the best to the data according to the AIC. This copula belongs to the Elliptical family and does not have any tail dependence, thus it is used when the strength of dependence is similar for all the values of the marginals. Furthermore, the association between the observed wind and simulated wind is relatively similar for all the values; see Fig. 4.7.

If we think about the wind in Portugal, and considering that Castelo Branco is situated in the interior centre of the country, strong winds are expected to occur mostly in Winter and weak winds in Spring and Summer. Winds in Autumn can be strong as well, and if we look at Tab. 4.17, we can see that the copula association parameter is the highest value obtained for the 3 fitted Gaussian Copulas. Moreover, the Gaussian parameter measures the strength of dependence between the variables and, as we can see in Tab. 4.18, the dependence between the observed and the simulated wind in Autumn is the second highest. These results reflect a closer agreement between the observed and the simulated data during the periods of higher wind speed. This is quite good, considering that the strongest winds are the ones which cause damages. We can also observe in Tab. 4.17 that none of the goodness-of-fit reject, at a significance level of 5%, the fitted copulas.

Tab. 4.17: Copula parameter estimates obtained through the MPLE method, coefficients of tail dependence and the p-values of the goodness-of-fit copula tests, $S_n^{(B)}$, S_n and $S_n^{(K)}$, for the copulas fitted to Autumn, Spring and Summer data of Castelo Branco

Season	Copula	$\hat{\theta}_{MPLE}$		Tail Dependence		p-values		
		$\hat{\rho}$	CI (95%)	λ_L	λ_U	$S_n^{(B)}$	S_n	$S_n^{(K)}$
Autumn	C_ρ^{gaus}	0.7538	(0.7018, 0.8058)	0	0	0.986	0.998	0.120
Spring	C_ρ^{gaus}	0.6008	(0.5165, 0.6851)	0	0	0.758	0.977	0.290
Summer	C_ρ^{gaus}	0.6006	(0.5255, 0.6769)	0	0	0.900	1	0.650

Tab. 4.18: Dependence measures of the data and of the fitted copulas for the station of Castelo Branco.

	Autumn		**Winter**		**Spring**		**Summer**	
	(X_A, Y_A)	$C_{A,\rho}^{gaus}$	(X_W, Y_W)	$C_{W,\alpha}^{gu}$	(X_{Sp}, Y_{Sp})	$C_{Sp,\rho}^{gaus}$	(X_{Su}, Y_{Su})	$C_{Su,\rho}^{gaus}$
ρ	0.7690	–	0.7850	–	0.6048	–	0.5751	–
τ	0.5499	0.5436	0.5767	0.5576	0.4166	0.4103	0.4159	0.4101
ρ_S	0.7338	0.7381	0.7614	0.7448	0.5938	0.5827	0.5886	0.5825

Tab. 4.19 shows the copula, its parameter estimates, its tail dependence coefficients and the p-values of the 3 goodness-of-fit tests performed for the remaining 8 stations.

Tab. 4.19: Copulas selected according to the AIC to fit the remaining selected stations. $\hat{\theta}_1$ and $\hat{\theta}_2$ are the copula parameters estimates obtained by MPLE.

Station	Season	Copula	$\hat{\theta}_{MPLE}$				Tail Dependence[1]		p-values		
			$\hat{\theta}_1$	CI (95%)	$\hat{\theta}_2$	CI (95%)	$\hat{\lambda}_L$	$\hat{\lambda}_U$	$S_n^{(B)}$	S_n	$S_n^{(K)}$
Aveiro	Autumn	C_α^{gu}	2.0422	(1.7814, 2.3030)	–	–	0	0.5959	0.950	1	0.560
	Winter	C_ρ^{gaus}	0.7847	(0.7344, 0.8350)	–	–	0	0	0.824	0.987	0.160
	Spring	C_α^{sc}	1.2066	(0.8532, 1.5600)	–	–	0	0.5630	0.191	0.114	0.300
	Summer	C_ρ^{gaus}	0.5396	(0.4522, 0.6270)	–	–	0	0	0.462	0.877	0.280
Bragança	Autumn	C_α^{gu}	2.0898	(1.8367, 2.3428)	–	–	0	0.6067	0.961	0.995	0.470
	Winter	C_α^{gu}	2.0245	(1.7790, 2.2500)	–	–	0	0.5917	0.999	1	0.560
	Spring	C_ρ^{guas}	0.5352	(0.4382, 0.6322)	–	–	0	0	0.971	1	0.780
	Summer	C_ρ^{gaus}	0.4539	(0.3597, 0.5482)	–	–	0	0	0.956	0.999	0.850
Coruche	Autumn	C_α^{gu}	2.3838	(2.1394, 2.6281)	–	–	0	0.6625	0.904	1	0.550
	Winter	C_ρ^{gaus}	0.8054	(0.7604, 0.8503)	–	–	0	0	0.982	1	0.160
	Spring	$C_{\rho\eta}^{t}$	0.5733	(0.4734, 0.6731)	6.7048	(*)	0.1877	0.1877	0.842	0.933	0.220
	Summer	C_α^{f}	3.2174	(2.3994, 4.0353)	–	–	0	0	0.934	1	0.830
Estremoz	Autumn	C_α^{gu}	2.2134	(1.9838, 2.4429)	–	–	0	0.6323	0.957	1	0.590
	Winter	C_α^{gu}	2.2304	(1.9566, 2.5043)	–	–	0	0.6355	0.988	0.975	0.770
	Spring	C_α^{gu}	1.6504	(1.4687, 1.8322)	–	–	0	0.4781	0.984	0.991	0.570
	Summer	C_α^{f}	4.7465	(3.7227, 5.7703)	–	–	0	0	0.971	1	0.600
Lisboa S1	Autumn	C_α^{gu}	2.4336	(2.0997, 2.7676)	–	–	0	0.6705	0.982	1	0.600
	Winter	C_ρ^{gaus}	0.8584	(0.8288, 0.8879)	–	–	0	0	0.965	1	0.950
	Spring	C_α^{gu}	1.7871	(1.5751, 1.9991)	–	–	0	0.5262	0.948	1	0.570
	Summer	C_α^{sc}	1.4206	(1.0555, 1.7857)	–	–	0	0.6139	0.196	0.037	0.980
Monção	Autumn	C_α^{gu}	1.6960	(1.4872, 1.9048)	–	–	0	0.4951	0.996	1	0.640
	Winter	C_α^{gu}	1.6001	(1.3942, 1.8060)	–	–	0	0.4578	0.973	1	0.370
	Spring	C_α^{gu}	1.4617	(1.3213, 1.6022)	–	–	0	0.3933	0.936	1	0.160
	Summer	C_α^{f}	2.2852	(1.4703, 3.1000)	–	–	0	0	0.906	1	0.330
Sines	Autumn	C_α^{gu}	2.1427	(1.8898, 2.3956)	–	–	0	0.6181	0.962	0.994	0.540
	Winter	C_α^{gu}	2.0819	(1.7861, 2.3776)	–	–	0	0.6049	0.970	1	0.610
	Spring	$C_{\rho\eta}^{t}$	0.7603	(0.7057, 0.8149)	7.0265	(*)	0.3263	0.3263	0.734	0.181	0.780
	Summer	C_α^{f}	6.5729	(5.5134, 7.6324)	–	–	0	0	0.769	1	0.450
Vila do Bispo	Autumn	C_α^{gu}	2.0844	(1.8111, 2.3657)	–	–	0	0.6064	0.986	1	0.680
	Winter	C_α^{gu}	2.2642	(1.9886, 2.5398)	–	–	0	0.6418	0.981	1	0.860
	Spring	C_ρ^{guas}	0.7934	(0.7884, 0.8385)	–	–	0	0	0.851	0.997	0.470
	Summer	C_ρ^{gaus}	0.8254	(0.7904, 0.8603)	–	–	0	0	0.941	1	0.040

(*) At present the asymptotic variance cannot be fully estimated if η is not fixed, thus it is not possible to provide a confidence interval; see the package *copula* manual. [1] the 0's are theoretical.

The comparison between copula estimation methods is presented for Castelo Branco's Winter in Tab. 4.20. Note that, for the full maximum likelihood method, MLE, new marginal parameters estimates were obtained. The copula parameter estimates vary slightly from the ones obtained by MPLE. We can see the estimation of the marginals together with the copula (MLE) or before the estimation of the copula (IFM), generates lower association parameters, while the rank-based estimators (MPLE, based on Kendall's tau or based on Spearman's rho) induce higher estimates of the copula parameter. The same conclusion is valid for the tail dependence coefficients. In this case, the upper tail coefficient is higher when we consider rank-based estimators rather than considering the marginal distributions. Regarding the estimation of the univariate distributions parameters, when estimated together with the copula parameter, they are also similar to the ones obtained before; see Tab. 4.7. Lastly, we obtain a mean relative error of about 5.55% in estimating C by the empirical copula C_n.

Tab. 4.20: Comparison between copula estimation methods for Castelo Branco's Winter season data. $\hat{\alpha}_C$ is the estimated copula parameter and $\hat{\mu}_X$, $\hat{\sigma}_X$, $\hat{\mu}_Y$ and $\hat{\sigma}_Y$ are the estimated parameters of the marginal distributions. The 95% confidence intervals are given between brackets. MRE is the mean relative error in estimating C by the empirical copula, C_n.

Winter

Method	$\hat{\theta}$					Tail Dep		MRE
	$\hat{\alpha}_C$	$\hat{\mu}_X$	$\hat{\sigma}_X$	$\hat{\mu}_Y$	$\hat{\sigma}_Y$	λ_L	$\hat{\lambda}_U$	(%)
MLE	2.2017	1.6664	0.4013	1.7859	0.3796	0	0.6300	–
	(1.9076,2.4958)	(1.6151,1.7176)	(0.3649,0.4377)	(1.7384,1.8334)	(0.3377,0.4034)	–	–	–
IFM	2.1389	–	–	–	–	0	0.6173	–
	(1.9133,2.3646)	–	–	–	–	–	–	–
$\tau^{(1)}$	2.3622	–	–	–	–	0	0.6590	–
	(2.0340,2.6905)	–	–	–	–	–	–	–
ρ_S	2.3435	–	–	–	–	0	0.6558	–
	(2.0093,2.6778)	–	–	–	–	–	–	–
C_n	–	–	–	–	–	–	–	5.55

$^{(1)}$ according to Tab.2.3.

Tab. 4.21 shows the comparison between the methods for copula estimation for the remaining seasons. The conclusion that the marginal distributions do not have a major influence on the estimation of the copula association parameter holds.

Tab. 4.21: Comparison between copula estimation methods for the remaining seasons of Castelo Branco.

Autumn

Method	$\hat{\theta}$					Tail Dep		MRE
	$\hat{\rho}_C$	$\hat{\alpha}_X$	$\hat{\beta}_X$	$\hat{\mu}_Y$	$\hat{\sigma}_Y$	λ_L	λ_U	(%)
MLE	0.7524	7.3557	1.349	1.7578	0.3733	0	0	-
	(0.6972,0.8076)	(6.0602,8.6512)	(1.1032,1.5948)	(1.7103,1.8053)	(0.3397,0.4069)	–	–	–
IFM	0.7529	–	–	–	–	0	0	–
	(0.7088,0.7969)	–	–	–	–	–	–	–
$\tau^{(1)}$	0.7603	–	–	–	–	0	0	–
	(0.6983,0.8222)	–	–	–	–	–	–	–
ρ_S	0.7497	–	–	–	–	0	0	–
	(0.6847,0.8148)	–	–	–	–	–	–	–
C_n	–	–	–	–	–	–	–	4.84

Spring

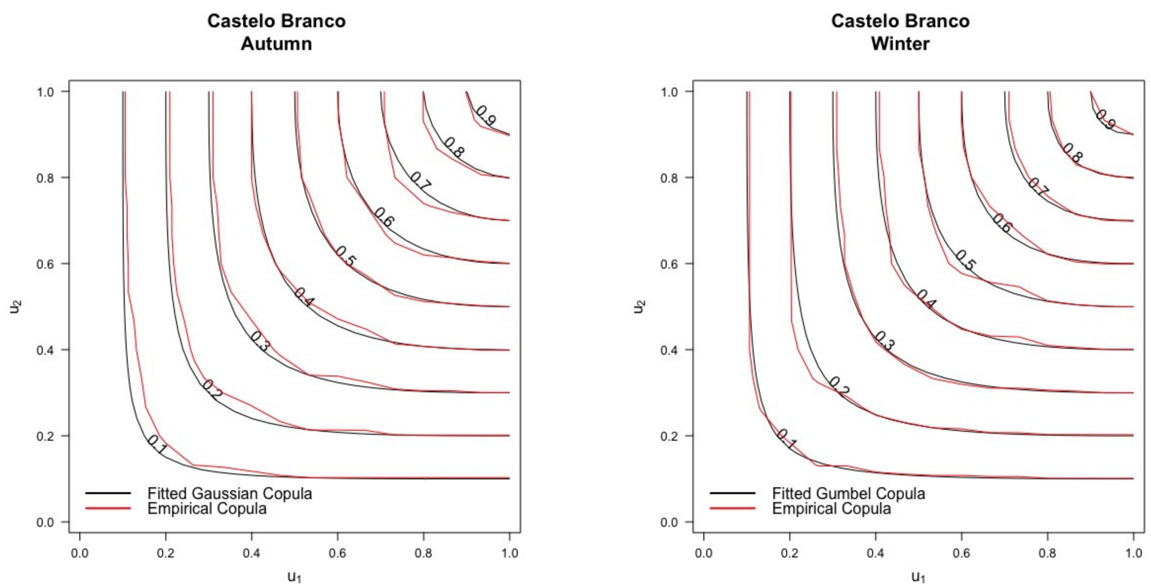
Meth.	$\hat{\theta}$						Tail Dep$^{(1)}$		MRE
	$\hat{\rho}_C$	$\hat{k}_X$	$\hat{c}_X$	$\hat{\lambda}_X$	$\hat{\alpha}_Y$	$\hat{\beta}_Y$	λ_L	λ_U	(%)
MLE	0.5854	1.6538	4.9681	0.1522	19.2348	2.4494	0	0	-
	(0.5029,0.6680)	(0.7638,2.5438)	(4.0973,5.8389)	(0.1274,0.1771)	(15.8321,22.6375)	(2.0104,2.8884)	–	–	–
IFM	0.5847	–	–	–	–	–	0	0	–
	(0.5128,0.6566)	–	–	–	–	–	–	–	–
$\tau^{(1)}$	0.6087	–	–	–	–	–	0	0	–
	(0.5274,0.6900)	–	–	–	–	–	–	–	–
ρ_S	0.6119	–	–	–	–	–	0	0	–
	(0.5308,0.6929)	–	–	–	–	–	–	–	–
C_n	–	–	–	–	–	–	–	–	4.26

Summer

Method	$\hat{\theta}$					Tail Dep$^{(1)}$		MRE
	$\hat{\rho}_C$	$\hat{\alpha}_X$	$\hat{\beta}_X$	$\hat{\omega}_Y$	$\hat{\delta}_Y$	λ_L	λ_U	(%)
MLE	0.5901	16.3913	2.7186	5.7371	8.6889	0	0	-
	(0.5059,0.6743)	(13.4575,19.3251)	(2.2244,3.2127)	(5.1702,6.3040)	(8.4855,8.8923)	–	–	–
IFM	0.5860	–	–	–	–	0	0	–
	(0.5136,0.6583)	–	–	–	–	–	–	–
$\tau^{(1)}$	0.6078	–	–	–	–	0	0	–
	(0.5256,0.6900)	–	–	–	–	–	–	–
ρ_S	0.6067	–	–	–	–	0	0	–
	(0.5224,0.6910)	–	–	–	–	–	–	–
C_n	–	–	–	–	–	–	–	7.4

$^{(1)}$ according to Tab.2.3.

In Fig. 4.8 we see the comparison between the contour plots of the fitted copula and the empirical one. Overall, the empirical copula, C_n, estimates well the fitted copula, especially in Autumn and Winter.



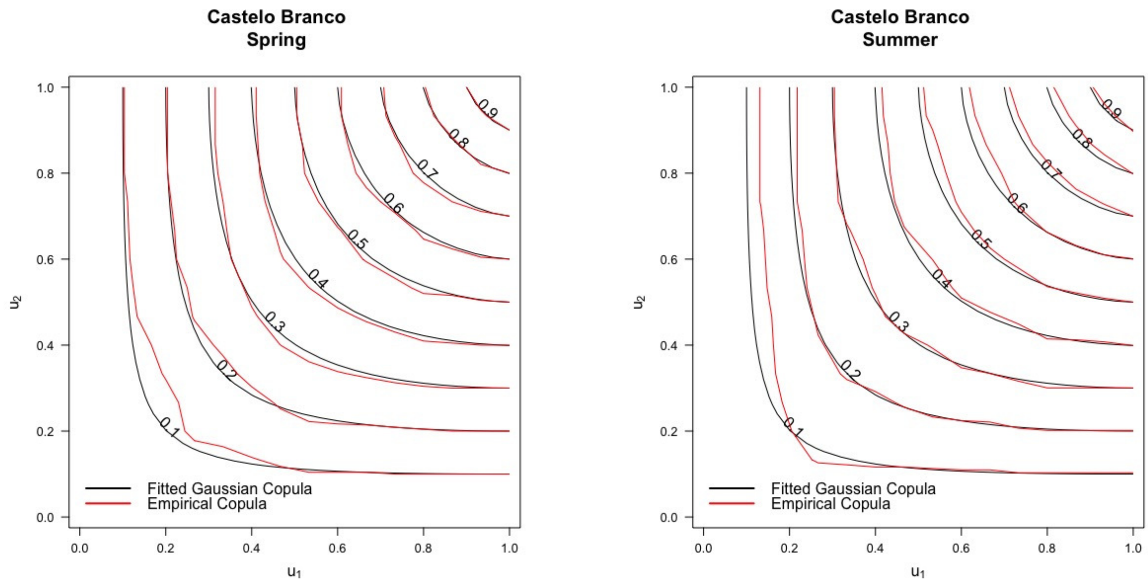


Fig. 4.8: Fitted copula vs Empirical copula of Castelo Branco.

Finally, we simulated 237, the sample size of Castelo Branco's Winter, realisations of a bivariate distribution constructed from the fitted Gumbel Copula and with the marginal distributions mentioned in the previous section, that is, $X \sim LN(1.6670, 0.3966)$ and $Y \sim LN(1.7819, 0.3642)$, and compared with the original data; see Fig. 4.9. We can see that the copula captures really well the wind speed of the joint model. Fig. 4.9 also shows the probability density function, $H(x,y)$ and its contour plot.

We also computed the probability of occurring strong winds [(a)] simultaneously in the observed and simulated data; see Tab. 4.22. Tab. 4.22 shows that, according to the copulas fitted to each of the seasons, it is in Spring that large observed data (at least 10 km/h) is more precisely simulated, that is, it has a higher value for $P[X > 10, Y > 10]$.

Tab. 4.22: Probabilities of occurring strong winds in Castelo Branco

	Autumn	Winter	Spring	Summer
$\overline{H}(10,10)$	0.0811	0.0895	0.1320	0.0951

In Fig. 4.10 we plotted the map of Portugal with all the copulas fitted to the 40 stations per season. We decided to code the copulas by colour and tail dependence characteristics. For instance, copulas without tail dependence are shaded in orange, copulas with lower tail dependence in red, with upper tail dependence in blue and with dependence in both tails in green. It is clear that in Autumn and Winter, the majority of copulas fitted have upper tail dependence, while there is more diversity in Spring and Summer. For instance, in both Autumn and Winter, 72.5% of the fitted copulas have upper tail dependence, while in Summer just only 20% show upper tail dependence. Moreover, in Summer, the majority of the fitted copulas, 52.5%, do not have any tail dependence; see Tab. 4.23.

Tab. 4.23: Percentage of tail dependence of the fitted copulas in the 40 stations per season.

Season	$\lambda_L = \lambda_U = 0$	$\lambda_L \neq 0 \wedge \lambda_U = 0$	$\lambda_L = 0 \wedge \lambda_U \neq 0$	$\lambda_L, \lambda_U \neq 0$
Autumn	20%	0%	72.5%	7.5%
Winter	25%	0%	72.5%	2.5%
Spring	30%	2.5%	40%	27.5%
Summer	52.5%	10%	20%	17.5%

[(a)] A wind is considered strong if its speed is greater than 10 km/h.

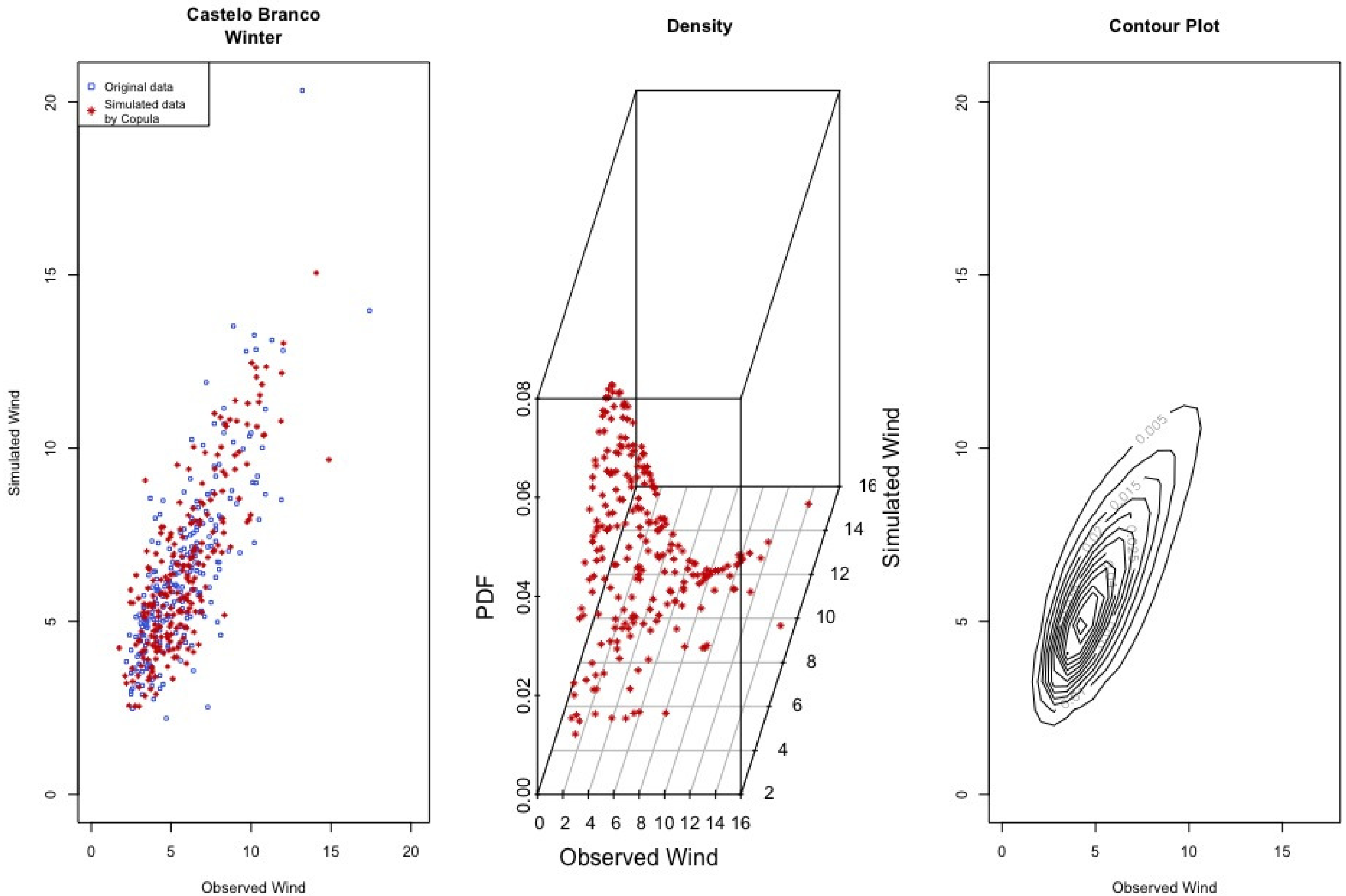


Fig. 4.9: Scatterplot of the original data and of the simulated bivariate distribution (left), probability density (middle) and contour plot (right) of the simulated bivariate distribution of Castelo Branco's Winter.

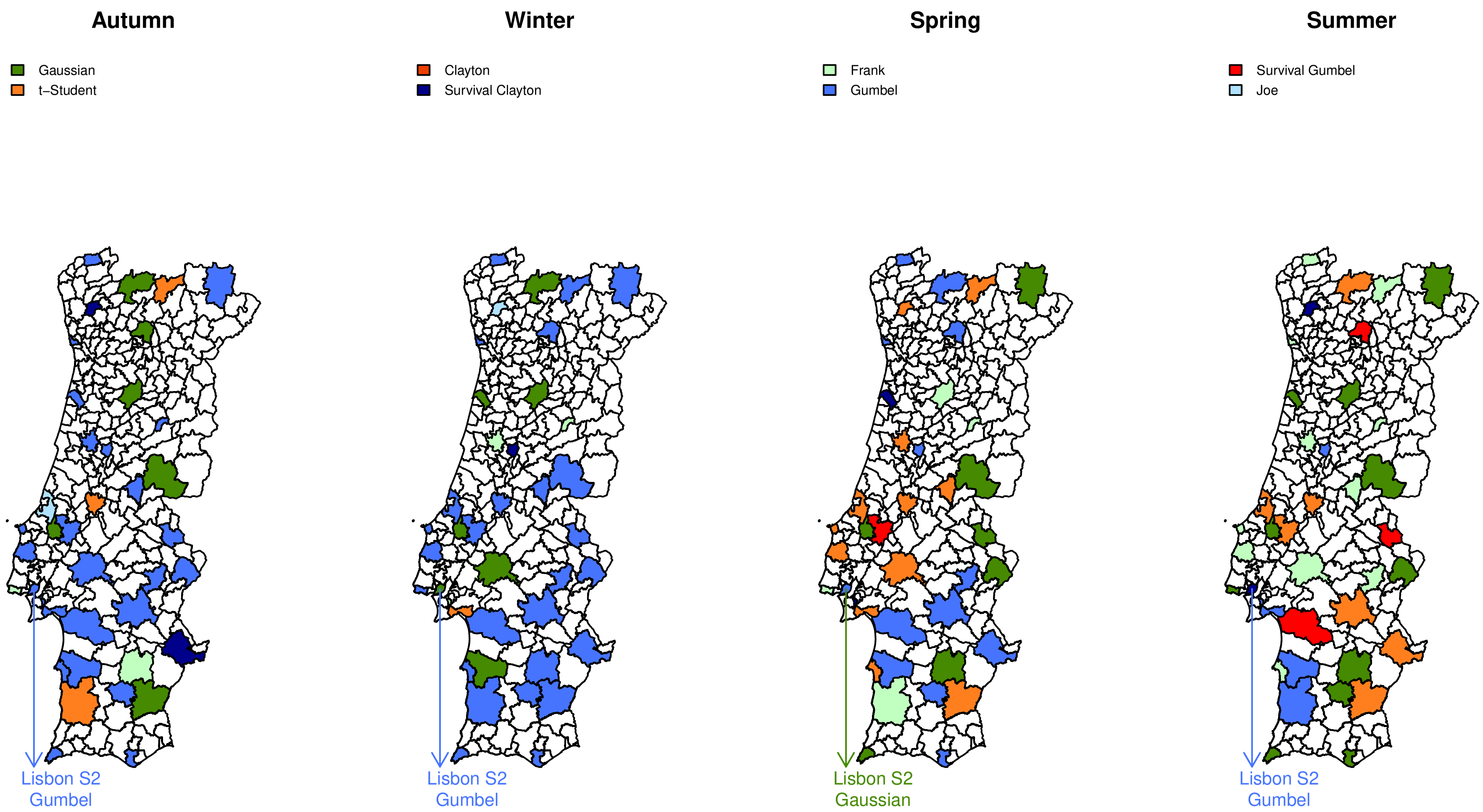


Fig. 4.10: Map of Portugal with the fitted copulas in the 40 stations per season. Copulas without tail dependence are in shades of green, upper tail dependence, in blue, with lower tail dependence, in red and with dependence on both tails, in orange.

We present in Tab. 4.24 the 10 meteorological stations and seasons which show the highest dependence between the observed wind speed and the simulated wind speed, as well as the 10 meteorological stations and seasons which have the lowest association between the variables. Although the copula parameter of a Gaussian or a Student *t*-Copula measures the strength of the dependence of the variables, we decided it would make more sense to present the rankings by a non-parametric concordance measure and we chose it to be the Kendall's tau, mainly because the relation between the copula parameter and Spearman's rho is not always as easy to obtain as the relation between the copula parameter and Kendall's tau. We can see that the highest dependences occur in Autumn and Winter and the lowest dependences in Summer. Therefore, in seasons when the wind speed registered is higher, the dependence between the observed data and the simulated by the simulator is also higher, which is most desirable, and in the season when the wind speed is lower, the dependence is also lower.

Tab. 4.24: The 10 stations with the highest Kendall's tau and the 10 stations with the lowest Kendall's tau.

Highest Kendall's tau			Lowest Kendall's tau		
Station	Season	τ	Station	Season	τ
Braga	Winter	0.6588	Lousã	Summer	0.1604
Lisboa S1	Winter	0.6571	Alcácer do Sal	Summer	0.1967
Moura	Spring	0.6404	Viseu	Summer	0.2279
Évora	Winter	0.6298	Viseu	Spring	0.2311
Cascais	Autumn	0.6254	Monção	Summer	0.2417
Santiago do Cacém	Winter	0.6247	Vila Real	Summer	0.2877
Vila do Bispo	Summer	0.6181	Alcobaça	Autumn	0.2940
Castro Verde	Winter	0.6094	Beja	Summer	0.2949
Coimbra	Autumn	0.6093	Bragança	Summer	0.2999
Évora	Autumn	0.6092	Alcobaça	Summer	0.3025

4.5 Bayesian Approach

We have only applied the Bayesian inference to the 9 stations selected, since we only want to compare different methods of estimation copula parameters and since the main scope of this thesis is to model the dependence between the observed wind speed with the simulated wind speed using copulas. As before, a complete analysis will be presented for Castelo Branco's Winter and the results for the other seasons and stations are shown at the end.

Following [Dos Santos Silva and Lopes, 2008], we jointly estimate all the models parameters. Recall that the marginal distributions fitted to the observed wind speed, X, and the simulated wind speed, Y, were Lognormal and the fitted copula was Gumbel. Therefore, the contribution of a single observation of X and Y to the logarithm of the likelihood is given by

$$\log(L(\mu_X,\sigma_X \mid x)) = -\log(x\sigma_X\sqrt{2\pi}) - \frac{(\log(x)-\mu_X)^2}{2\sigma_X^2}, \tag{4.2}$$

$$\log(L(\mu_Y,\sigma_Y \mid y)) = -\log(y\sigma_Y\sqrt{2\pi}) - \frac{(\log(y)-\mu_Y)^2}{2\sigma_Y^2}, \tag{4.3}$$

respectively, and, recalling (2.69),

$$\log(L(\alpha_C \mid u,v)) = -\log(uv) + (\alpha_C-1)\log(\log(u)\log(v)) + \log[w^{\frac{2}{\alpha_C}-2} + (\alpha_C-1)w^{\frac{1}{\alpha_C}-2}] - w^{\frac{1}{\alpha_C}}, \tag{4.4}$$

where $w = (-\log(u))^{\alpha_C} + (-\log(v))^{\alpha_C}$. Thus, (3.14) is straightforward expressed as

$$\log(L(\mu_X,\sigma_X,\mu_Y,\sigma_Y,\alpha_C \mid x,y)) = \log(L(\mu_X,\sigma_X \mid x)) + \log(L(\mu_Y,\sigma_Y \mid y)) + \log(L(\alpha_C \mid u,v)). \tag{4.5}$$

For the implementation in JAGS (and in WinBUGS), the log-likelihood is defined for a single observation at a time in a cycle of all pairs (x_i, y_i), $i = 1, \ldots, n$; see Appendix C for the code. Non-informative priors were used for all the model parameters; see [Dos Santos Silva and Lopes, 2008]. A Gamma distribution with mean 1 and variance 10^3 is used for σ_X and σ_Y and a Normal distribution with mean 0 and precision 10^{-4} for μ_X and μ_Y. The choice of the non-informative priors for the copula parameters are based mainly on their relationship with Kendall's tau (recall Tab. 2.3) and, in the case of Elliptical copulas, with Pearson's correlation coefficient. For the Gumbel Copula, the relation is $\alpha_C = \frac{1}{1-\tau}$. If we take $\theta = \dfrac{1}{\alpha_C}$, we get $\tau = 1 - \theta$. Moreover, $\tau \to 1$ implies $\theta \to 0$. For this reason a Beta distribution with both shape and scale parameters equal to $\frac{1}{2}$ is used as a prior for θ. Note that $\theta \in (0, 1]$; see [Shemyakin and Kniazev, 2017].

All the computations were performed with JAGS software, with *runjags* package and the function *autorun.jags*, which assesses automatically for chain convergence. It then runs a number of iterations for achieving the necessary sample size length according to Raftery and Lewis's diagnosis. However, since it uses BUGS language, it is easily adapted to WinBUGS. The "zero tricks" described in the WinBUGS user manual suffers a slightly modification when adapted to JAGS. Nonetheless, it was used to encode the log-likelihood and assure that the correct likelihood is obtained.

Two Markov Chains, starting from different initial values, were used and convergence as well as the desired sample length were reached after 4000 iterations. A burn-in of 5000 iterations was also used. Tab. 4.25 summarises the posterior results and Fig. 4.11 shows the trace plots, the histograms and the autocorrelation plots of the results. Fig. 4.12 shows the comparison between the 95% confidence intervals obtained with the classical maximum likelihood approach and the 95% credible intervals. One can see that, although the parameters estimates vary little from the ones obtained in the classical approach, the credible intervals are narrowed, which constitutes an advantage of using Bayesian inference to estimate the parameters, even with non-informative priors.

Tab. 4.25: Bayesian estimation for Castelo Branco's Winter.

Winter

	Mean	SD	MCE	CI (95%)
μ_X	1.6663	0.01691	0.000328	(1.6347, 1.7000)
σ_X	0.3989	0.01240	0.000230	(0.3746, 0.4229)
μ_Y	1.7847	0.01533	0.000295	(1.7546, 1.8138)
σ_Y	0.3679	0.01119	0.000197	(0.3463, 0.3909)
α_C	2.1771	0.12844	0.002395	(1.9389, 2.4418)

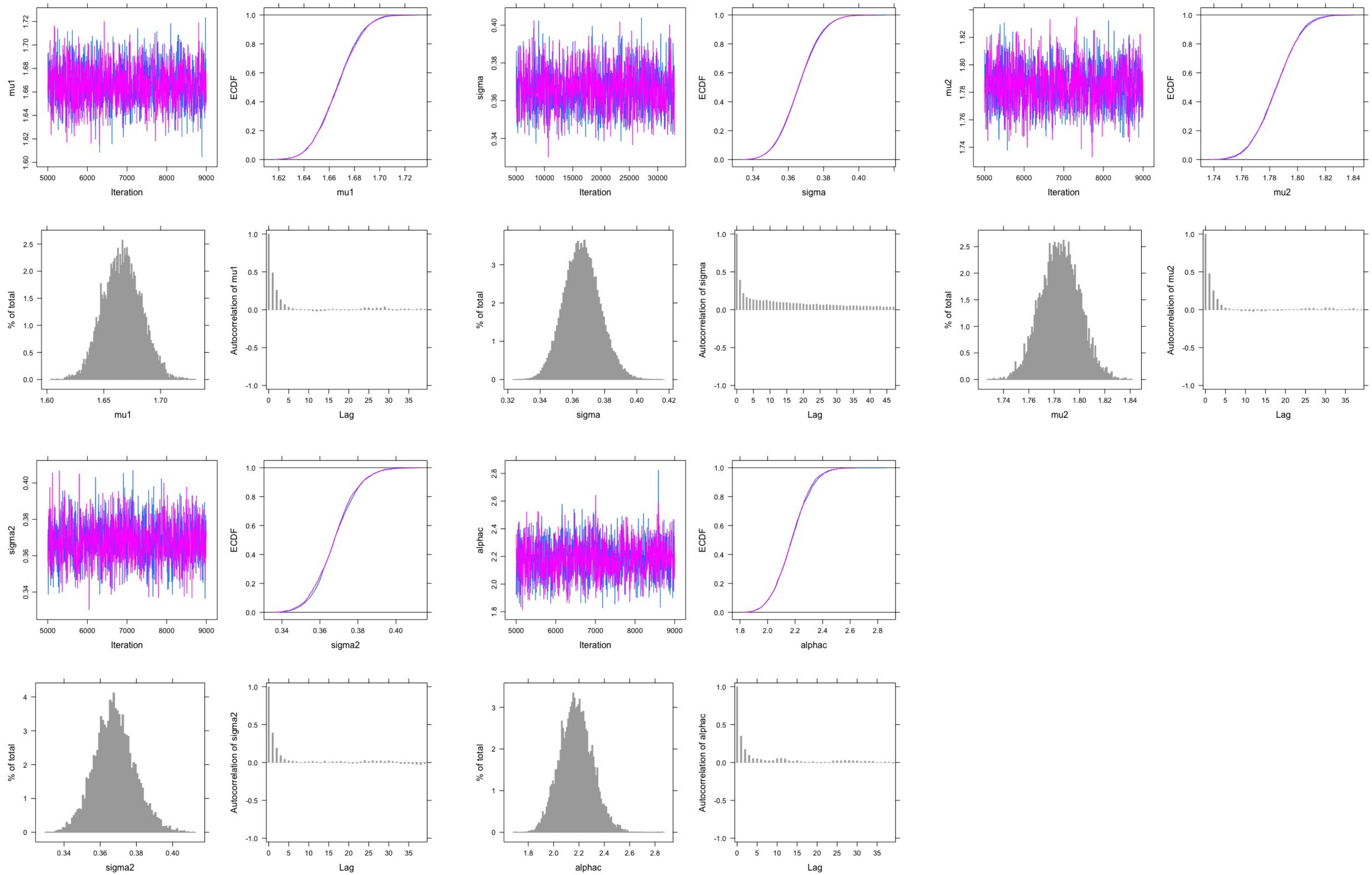


Fig. 4.11: Plots of the posterior results of Castelo Branco’s Winter.

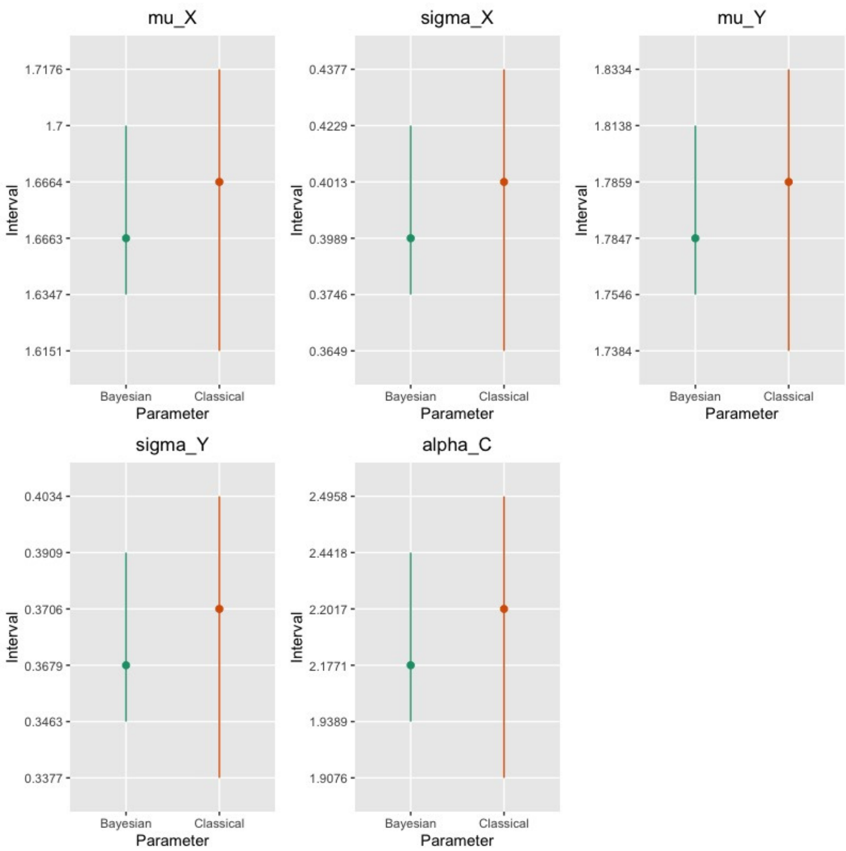


Fig. 4.12: Comparison of parameter estimation based on MLE classical method and MLE bayesian method through 95% credible and 95% confidence intervals for μ_X (upper left), σ_X (upper middle), μ_Y (upper right), σ_Y (lower left), α_C (lower right) for Castelo Branco's Winter.

The posterior results for the remaining seasons of Castelo Branco are shown in Tab. 4.26 and the comparison between the confidence and the credible intervals for each parameter in Fig. 4.13. The difference between the parameters estimates remains little, but the Bayesian inference provides a narrowed range for the parameters value. For the remaining 8 stations; see Tab. 4.27 and 4.28.

Tab. 4.26: Bayesian estimation for the remaining seasons of Castelo Branco.

Autumn					**Spring**				
	Mean	SD	MCE	CI (95%)		Mean	SD	MCE	CI (95%)
α_X	7.3397	0.44565	0.007675	(6.5328,8.2843)	k_X	1.5797	0.33597	0.006285	(0.9937,2.2286)
β_X	1.3462	0.08464	0.00146	(1.1868,1.5149)	c_X	5.0540	0.33618	0.005996	(4.4008,5.7034)
–	–	–	–	–	λ_X	0.1554	0.00952	0.000189	(0.1370,0.1736)
μ_Y	1.7575	0.01575	0.000169	(1.7273,1.7879)	α_Y	19.2451	1.22810	0.031456	(16.9061,21.6251)
σ_Y	0.3745	0.11580	0.000135	(0.3521,0.3976)	β_Y	2.4505	0.15836	0.004060	(2.1503,2.7593)
ρ	0.7475	0.02545	0.000294	(0.6969,0.7959)	ρ	0.5785	0.04016	0.000506	(0.4987,0.6555)

Summer				
	Mean	SD	MCE	CI (95%)
α_X	16.5443	1.04700	0.016632	(14.5952,18.7066)
β_X	2.7436	0.17652	0.002802	(2.4005,3.0938)
ω_Y	5.7618	0.19826	0.004746	(5.3792,6.1521)
δ_Y	8.6897	0.06956	0.000717	(8.5546,8.8252)
ρ	0.5806	0.03996	0.000376	(0.4989,0.6540)

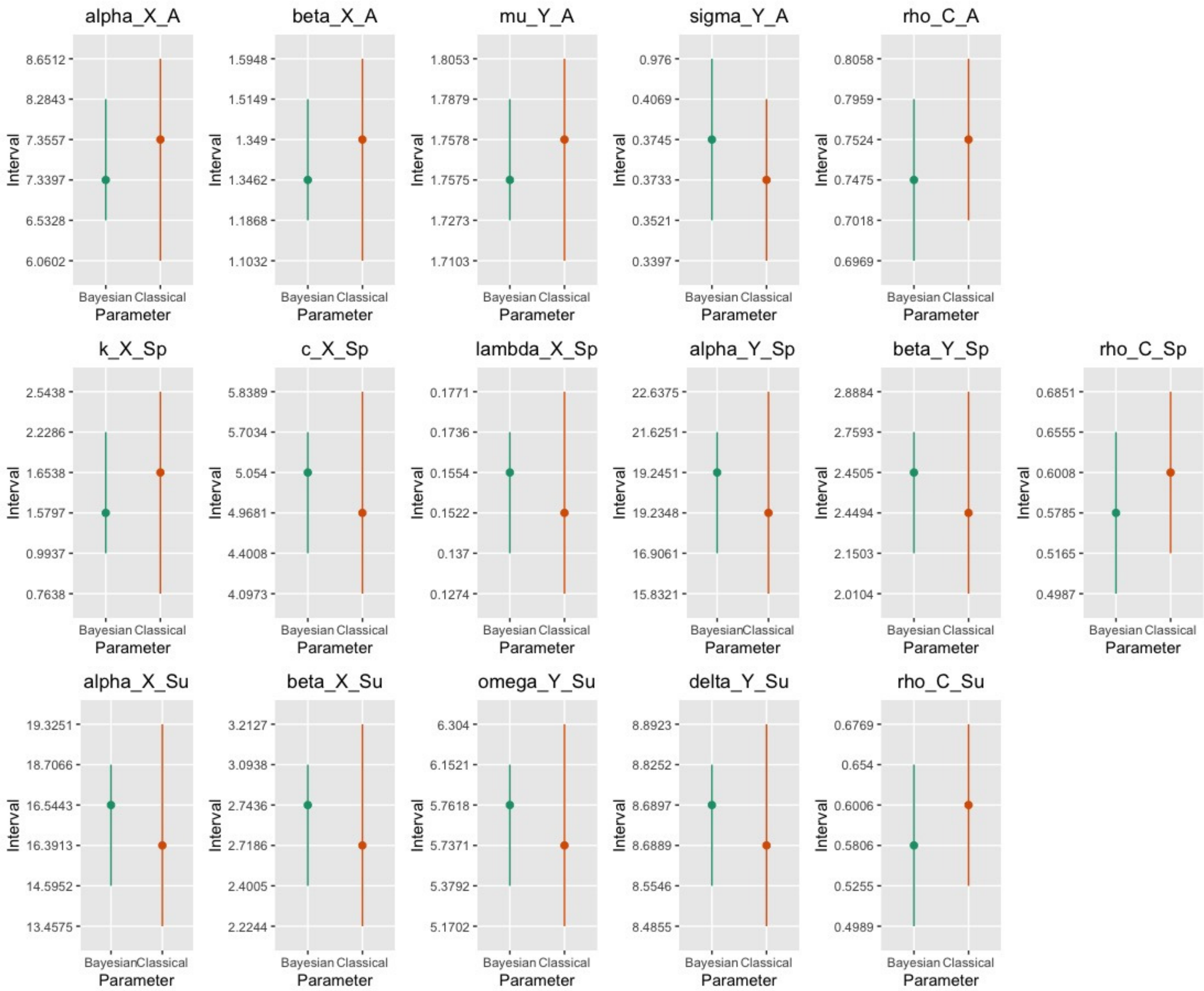


Fig. 4.13: Comparison of parameter estimation based on MLE classical method and MLE bayesian method through 95% credible and 95% confidence intervals for the remaining seasons of Castelo Branco.

Tab. 4.27: Bayesian estimation for the remaining stations selected in Autumn and Winter

	Autumn					Winter				
		Mean	SD	MCE	CI (95%)		Mean	SD	MCE	CI (95%)
Aveiro	μ_X	1.6184	0.01677	0.000303	(1.5861, 1.6510)	μ_X	1.7039	0.01934	0.000367	(1.6675, 1.7424)
	σ_X	0.3935	0.01203	0.000206	(0.3705, 0.4175)	σ_X	0.4648	0.01434	0.000248	(0.4374, 0.4931)
	μ_Y	1.8716	0.01498	0.000268	(1.8411, 1.8998)	μ_Y	1.9386	0.01700	0.000328	(1.9064, 1.9723)
	σ_Y	0.3616	0.01112	0.000203	(0.3400, 0.3830)	σ_Y	0.4034	0.01254	0.000220	(0.3793, 0.4287)
	α_C	1.9862	0.11604	0.001917	(1.7728, 2.2205)	ρ_C	0.7763	0.02296	0.000385	(0.7318, 0.8220)
Bragança	μ_X	1.5656	0.01982	0.000214	(1.5269, 1.6050)	α_X	4.7656	0.29070	0.004783	(4.2079, 5.3424)
	σ_X	0.4652	0.01408	0.000169	(0.4379, 0.4929)	β_X	0.8694	0.05548	0.000918	(0.7606, 0.9779)
	α_Y	6.3619	0.33722	0.005331	(5.0458, 6.3619)	α_Y	5.0361	0.30717	0.005073	(4.4373, 5.6374)
	β_Y	0.9405	0.05758	0.000916	(0.8275, 1.0539)	β_Y	0.8494	0.05377	0.000895	(0.7455, 0.9536)
	α_C	2.0519	0.11826	0.001419	(1.8279, 2.2914)	α_C	1.9662	0.11474	0.001464	(1.7447, 2.1899)
Coruche	α_X	5.1022	0.32263	0.005659	(4.4789, 5.7246)	k_X	5.7607	3.87570	0.191800	(1.9442, 12.134)
	β_X	0.9756	0.06381	0.001125	(0.8555, 1.1024)	c_X	2.9132	0.17088	0.004853	(2.5788, 3.2414)
	–	–	–	–	–	λ_X	0.1072	0.02063	0.000755	(0.0669, 0.1477)
	α_Y	6.3129	0.38991	0.007971	(5.5563, 7.0794)	μ_Y	1.8514	0.01566	0.000111	(1.8205, 1.8815)
	β_Y	0.9334	0.05924	0.001102	(0.8230, 1.0535)	σ_Y	0.3659	0.01164	0.000082	(0.3435, 0.3892)
	α_C	2.3290	0.14610	0.001850	(2.0413, 2.6098)	ρ_C	0.7896	0.02216	0.000163	(0.7454, 0.8317)
Estremoz	μ_X	1.5302	0.01479	0.000288	(1.5020, 1.5604)	α_X	6.7512	0.42210	0.008426	(5.9548, 7.6040)
	σ_X	0.3556	0.01086	0.000197	(0.3346, 0.3765)	β_X	1.3554	0.08643	0.001724	(1.1902, 1.5266)
	μ_Y	1.7153	0.01538	0.000292	(1.6864, 1.7454)	μ_Y	1.7462	0.01607	0.000164	(1.7143, 1.7774)
	σ_Y	0.3751	0.01149	0.000211	(0.3542, 0.3996)	σ_Y	0.3650	0.01141	0.000145	(0.3431, 0.3878)
	α_C	2.1993	0.13063	0.002379	(1.9529, 2.4589)	α_C	2.2016	0.13878	0.001614	(1.9323, 2.4747)
Lisboa S1	k_X	1.0005	0.15349	0.002398	(0.7213, 1.3063)	k_X	1.4982	0.29140	0.005761	(1.0078, 2.0757)
	c_X	5.2935	0.36900	0.005219	(4.5798, 6.0195)	c_X	4.3566	0.27930	0.004845	(3.8118, 4.9169)
	λ_X	0.1699	0.00768	0.000116	(0.1550, 0.1849)	λ_X	0.1502	0.00976	0.000186	(0.1318, 0.1693)
	μ_Y	1.8260	0.01450	0.000140	(1.7975, 1.8542)	μ_Y	1.8648	0.01408	0.000158	(1.8385, 1.8935)
	σ_Y	0.3401	0.01072	0.000104	(0.3204, 0.3626)	σ_Y	0.3341	0.01024	0.000116	(0.3141, 0.3539)
	α_C	2.3989	0.14835	0.001408	(2.1081, 2.6872)	ρ_C	0.8386	0.01694	0.000189	(0.8066, 0.8718)
Monção	k_X	3.3776	1.14870	0.020981	(1.7155, 5.5854)	k_X	3.0491	0.91190	0.015743	(1.6628, 4.7806)
	c_X	3.5393	0.21415	0.003018	(3.1269, 3.9583)	c_X	3.5853	0.21790	0.002960	(3.1666, 4.0174)
	λ_X	0.1341	0.01547	0.000260	(0.1035, 0.1638)	λ_X	0.1369	0.01428	0.000233	(0.1040, 0.1537)
	μ_Y	2.0054	0.01439	0.000102	(1.9762, 2.0328)	μ_Y	2.0383	0.01428	0.000115	(2.0100, 2.0657)
	σ_Y	0.3224	0.01034	0.000074	(0.3019, 0.3425)	σ_Y	0.3096	0.01042	0.000083	(0.2900, 0.3305)
	α_C	1.6387	0.09382	0.000669	(1.4555, 1.8201)	α_C	1.5483	0.08852	0.000706	(1.3857, 1.7303)
Sines	μ_X	1.9079	0.01596	0.000324	(1.8785, 1.9398)	μ_X	1.9078	0.01540	0.000299	(1.8777, 1.9379)
	σ_X	0.3740	0.01138	0.000219	(0.3519, 0.3967)	σ_X	0.3549	0.01068	0.000191	(0.3336, 0.3751)
	μ_Y	2.0102	0.01384	0.000277	(1.9832, 2.0370)	μ_Y	2.0594	0.01376	0.000258	(2.0310, 2.0858)
	σ_Y	0.3311	0.01020	0.000193	(0.3114, 0.3510)	σ_Y	0.3236	0.00988	0.000182	(0.3046, 0.3432)
	α_C	2.1360	0.12560	0.002220	(1.8874, 2.3796)	α_C	2.0285	0.12053	0.002111	(1.8000, 2.2693)
Vila do Bispo	α_X	7.4629	0.45780	0.008442	(6.5689, 8.3475)	ω_X	2.4462	0.08361	0.001609	(2.2857, 2.6090)
	β_X	1.0643	0.06630	0.001195	(0.9310, 1.1912)	δ_X	8.7028	0.15729	0.001986	(8.3885, 9.0099)
	μ_Y	2.0766	0.01470	0.000164	(2.0495, 2.1066)	μ_Y	2.1733	0.01678	0.000202	(2.1394, 2.2059)
	σ_Y	0.3495	0.01079	0.000133	(0.3277, 0.3697)	σ_Y	0.3926	0.01351	0.000162	(0.3666, 0.4190)
	α_C	2.0531	0.11803	0.001469	(1.8253, 2.2813)	α_C	2.6702	0.18375	0.001976	(2.2982, 3.0169)

Tab. 4.28: Bayesian estimation for the remaining stations selected in Spring and Summer.

	Spring					Summer				
		Mean	SD	MCE	CI (95%)		Mean	SD	MCE	CI (95%)
Aveiro	μ_X	1.8311	0.01338	0.000221	(1.8052, 1.8573)	k_X	0.9028	0.17650	0.002927	(0.5896, 1.2605)
	σ_X	0.2998	0.00934	0.000151	(0.2820, 0.3183)	c_X	8.1053	0.65850	0.010124	(6.8895, 9.4296)
	–	–	–	–	–	λ_X	0.1708	0.00638	0.000105	(0.1573, 0.1825)
	μ_Y	2.1000	0.01219	0.000213	(2.0775, 2.1250)	μ_Y	2.1096	0.01187	0.000133	(2.0851, 2.1320)
	σ_Y	0.2755	0.00855	0.000141	(0.2594, 0.2924)	σ_Y	0.2727	0.00892	0.000100	(0.2557, 0.2902)
	α_C	1.1056	0.14680	0.000249	(0.8329, 1.4034)	ρ_C	0.5397	0.04379	0.000506	(0.4549, 0.6233)
Bragança	μ_X	1.7653	0.01170	0.000083	(1.7429, 1.7881)	k_X	1.2901	0.25232	0.003768	(0.8642, 1.8077)
	σ_X	0.2646	0.00858	0.000061	(0.2485, 0.2821)	c_X	6.3895	0.43220	0.005863	(5.575 0, 7.2471)
	k_Y	8.3453	16.18183	1.048122	(1.3564, 24.7921)	λ_X	0.1750	0.00805	0.000120	(0.1592, 0.1906)
	c_Y	4.6278	0.33690	0.006385	(3.9611, 5.2560)	ω_Y	4.7901	0.16875	0.004261	(4.4582, 5.1248)
	λ_Y	0.0922	0.01683	0.000597	(0.0549, 0.1223)	δ_Y	7.3788	0.07212	0.000854	(7.2396, 7.5196)
	ρ_C	0.5255	0.04398	0.000328	(0.4278, 0.6074)	ρ_C	0.4591	0.04937	0.000519	(0.3513, 0.5437)
Coruche	k_X	3.2461	1.22760	0.021743	(1.6030, 5.6219)	k_X	2.2596	0.47588	0.004472	(1.4125, 3.1722)
	c_X	5.7610	0.35542	0.004999	(5.0778, 6.4619)	c_X	8.5211	0.48343	0.004083	(7.5355, 9.4314)
	λ_X	0.1246	0.00953	0.000156	(0.1051, 0.1421)	λ_X	0.1495	0.00515	0.000048	(0.1395, 0.1595)
	k_Y	3.3380	1.08420	0.016684	(1.7033, 5.4978)	k_Y	9.1976	8.94360	0.619330	(1.7673, 26.4310)
	c_Y	7.9015	0.45302	0.005419	(6.9948, 8.7650)	c_Y	7.9756	0.50202	0.016357	(7.0626, 8.9637)
	λ_Y	0.0947	0.00484	0.000070	(0.0850, 0.1040)	λ_Y	0.0857	0.00878	0.000419	(0.0672, 0.1007)
	ρ_C	0.5530	0.04735	0.000343	(0.4617, 0.6462)	α_C	3.3048	0.45732	0.003258	(2.4390, 4.2299)
	η_C	13.3077	83.85160	0.592921	(2.5629, 27.8197)	–	–	–	–	–
Estremoz	α_X	14.3298	0.88831	0.014105	(12.5628, 16.0117)	μ_X	1.5859	0.00892	0.000079	(1.5682, 1.6032)
	β_X	2.7467	0.17203	0.002742	(2.3963, 3.0683)	σ_X	0.2093	0.00661	0.000056	(0.1970, 0.2229)
	ω_Y	4.8550	0.16262	0.003736	(4.5482, 5.1853)	ω_Y	6.3249	0.22967	0.005826	(5.8801, 6.7703)
	δ_Y	7.8749	0.07312	0.000880	(7.7385, 8.0237)	δ_Y	7.6124	0.05418	0.000512	(7.5095, 7.7220)
	α_C	1.6134	0.08994	0.000944	(1.4414, 1.7942)	α_C	4.6197	0.46830	0.003939	(3.6991, 5.5328)
Lisboa S1	μ_X	1.8612	0.00903	0.000102	(1.8424, 1.8787)	α_X	25.8421	1.61027	0.032470	(22.7319, 29.0276)
	σ_X	0.2120	0.00655	0.000081	(0.1995, 0.2254)	β_X	3.7625	0.23400	0.004765	(3.3375, 4.2497)
	α_Y	21.302	1.28864	0.029771	(18.7913, 23.7887)	α_Y	22.5375	1.40753	0.028032	(19.8460, 25.3190)
	β_Y	2.6660	0.16150	0.003746	(2.3483, 2.9744)	β_Y	2.7339	0.17094	0.003344	(2.4147, 3.0790)
	α_C	1.7804	0.09931	0.001294	(1.5768, 1.9704)	α_C	1.3624	0.16399	0.001831	(1.0344, 1.6802)
Monção	μ_X	1.6706	0.01314	0.000201	(1.6448, 1.6964)	μ_X	1.6765	0.01031	0.000144	(1.6564, 1.6965)
	σ_X	0.3008	0.00956	0.000145	(0.2821, 0.3194)	σ_X	0.2330	0.00763	0.000107	(0.2190, 0.2488)
	μ_Y	2.0708	0.01000	0.000155	(2.0509, 2.0898)	μ_Y	2.0289	0.00834	0.000124	(2.0121, 2.0449)
	σ_Y	0.2282	0.00745	0.000117	(0.2136, 0.2427)	σ_Y	0.1835	0.00611	0.000088	(0.1715, 0.1954)
	α_C	1.4952	0.08336	0.001280	(1.3409, 1.6638)	α_C	2.1881	0.40916	0.005935	(1.4190, 3.0155)
Sines	α_X	13.8450	0.84217	0.017118	(12.2148, 15.4888)	ω_X	4.5040	0.15478	0.003870	(4.2049, 4.8150)
	β_X	1.8179	0.11296	0.002228	(1.6061, 2.0478)	δ_X	8.7685	0.08245	0.000915	(8.6026, 8.0280)
	α_Y	21.8748	1.35965	0.021832	(19.2397, 24.5667)	α_Y	25.2443	1.61920	0.040804	(22.1373, 28.4041)
	β_Y	2.4638	0.15495	0.003611	(2.1652, 2.7719)	β_Y	2.9231	0.19110	0.004800	(2.5668, 3.3030)
	ρ_C	9.7484	0.03159	0.000370	(0.6861, 0.8083)	α_C	6.6551	0.56447	0.006379	(5.5647, 7.7448)
	η_C	5.4138	5.23024	0.049810	(2.0361, 10.1294)	–	–	–	–	–
Vila do Bispo	α_X	4.1034	0.13410	0.001889	(3.8313, 4.3549)	ω_X	4.1017	0.13900	0.003009	(3.8404, 4.3852)
	β_X	8.7198	0.09279	0.000781	(8.5463, 8.9084)	δ_X	9.5853	0.10090	0.001147	(9.3873, 9.7871)
	k_Y	4.0789	1.89441	0.045875	(1.8147, 7.0444)	ω_Y	5.2524	0.16883	0.004182	(4.9280, 5.5879)
	c_Y	5.5371	9.31268	0.004910	(4.9375, 6.1604)	δ_Y	10.2182	0.08436	0.001004	(10.0604, 10.3924)
	λ_Y	0.0817	0.00673	0.000127	(0.0679, 0.0939)	–	–	–	–	–
	ρ_C	0.7902	0.02115	0.000184	(0.7456, 0.8284)	ρ_C	0.8154	0.01897	0.000233	(0.7786, 0.8522)

5 | Comments, Conclusions and Future Work

The main purpose of this work was to model the dependence between real wind speed data and simulated wind speed data using bivariate copulas. The major benefit of using the simulated wind speed data towards the real one is the lack of missing values. Therefore, studying the dependence between these two variables is of an extreme importance in order to improve the simulated data, i.e, to bring the simulated wind speeds in line with the observed wind speeds. If the simulated wind data matches the daily maximum wind speeds in all the domain, that means that the model used to simulate the data is suitable. Moreover, one can predict the wind values for a certain region and avoid possible damages or company losses. The problem is that the data provided by the simulator tend to differ from the observed data, especially on the tails. Beyond allowing to model the joint distribution separately from the marginal distributions, the tail dependence property is extremely attractive for addressing the problem mentioned above. The knowledge of the dependence between the extreme values would help to improve the model used to simulate the wind data and consequently achieve more similarities with the real data, especially in the tails. The original data consists in daily maximum wind speeds, measured in km/h, observed in 117 meteorological stations spread out in the continental part of Portugal from 1997 to 2013 and simulated wind speeds produced by a simulator at a regular grid of 81 km^2 grid cell size. However, the data registered had an extremely high proportion of missing observations, which reached 90% in some stations. Thus, we have just considered only those with less that 30% of NAs, resulting in 40 stations out of 117. The missing values were removed as well as the wind speeds equal to 0. Considering that the years from 1997 to 1999 showed large periods without any observation recorded, and 2013 had just observations until February, the analysis was made only from 2000 to 2012. We then divided the data set in 4 parts, each of which represents a season, since is natural to expect that the wind speed has different behaviours in the 4 seasons. Moreover, the analysis performed has supported this decision. Finally, we observed the presence of short-term dependence within the series of observed wind speed and within the series of simulated wind speed. Therefore, in order to overcome this problem, we have retained every 5 observations, which proved to be indeed the right choice.

In Chapter 2, we highlight the key definitions in order to give to the reader a better understanding of the background of Copula Theory. Moreover, the most common families of copulas are mentioned and their characteristics described. With respect to the Elliptical family, the characteristics of the Gaussian and Student *t*-Copulas were summarised, while for the Archimedean family, we look with detail into the Clayton, Frank and Gumbel Copulas. Chapter 3 is divided in 2 parts: we first present several methods of estimation of copulas, and then we address the selection of the right copula model as well as the goodness-of-fit of the marginal distributions to the data. We consider these two chapters essential to the understanding of the procedures used in Chapter 4.

Chapter 4 presents the methodology used and the results obtained when we modelled the daily maximum wind speeds and the wind speeds produced by the simulator with copulas. At first we fit marginal distributions to the data. In the literature, it is common to use the Weibull distribution to model wind type data; see [Mert and Karakus, 2015], [Pobocikova *et al*, 2017], [Shepherd, 1978], [Harris and Cook, 2014]. However, one should not fit this distribution without any statistical support.

Moreover, we found the Gamma and the Lognormal distributions to be more suitable to deal with our data. In fact, these distributions were fitted in 31.25% and 30.31%, respectively, in all cases, while Weibull modelled the variables only in 10.6%. When assessing the fit of the marginal distributions to the variables, we also took into consideration the behaviour on the right tail. Recall that it is important to study extreme values of wind speed, because strong winds are the ones which cause the most relevant damage. After adjusting the distributions to the variables, we found some cases where the simulated wind speed data seemed to have been shifted to the right towards the observed wind speed. In particular, this occurred in Spring and Summer, when the wind is milder. It would be important, in the future, to improve the model which produces the simulated wind speeds, in order to obtain a better match between the variables.

We then searched for the best copula model to capture the dependence of the variables. We used the AIC to selected the copula among 40 families of copulas. However, we were not able to simulate a bivariate distribution of some of the copulas. Thus, we restricted the selection to 6 families and the corresponding survival copulas: Gaussian, Student *t*, Clayton and Survival Clayton, Frank, Gumbel and Survival Gumbel and, occasionally, Joe. We decided to estimate the copula parameter by maximum pseudo-likelihood method, acknowledging [Genest and Favre, 2007] where it is stated that the dependence structure captured by the copula is not affected by the choice of the marginal distributions. A comparison between all possible methods mentioned in Section 3.1 was provided only for the station of Castelo Branco. All the estimation methods showed quite similar results. However, in most cases, the influence of the marginal distributions lowered both the values of the parameters and the tail dependence coefficients (when they exist), either by estimating them by fully maximum likelihood or by the *"Inference Functions for Margins"* method.

The Gumbel Copula was fitted in 45% of the meteorological stations, which is characterised to have upper tail dependence. Moreover, 65% of the adjusted copulas have upper tail dependence and modelled especially the Autumn and Winter seasons, while just about 32% of the cases were modelled by a copula without any type of tail dependence, mostly in Summer. Therefore, in Autumn and Winter, high values of daily maximum wind speed and high values of simulated wind speed are associated, which means that the simulated data seems to match really well the observed data for strong winds but not so well for weak winds.

At last, we applied Bayesian inference in 9 selected meteorological stations: Aveiro, Bragança, Castelo Branco, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo. We chose to jointly estimate all parameters in order to obtain a complete characterisation of the posterior distribution. We used non-informative priors and, as stated by [Joe, 2014], the likelihood dominated the prior distribution and similar estimates were obtained. However, Bayesian inference provided narrower intervals for the parameters values.

As mentioned in the beginning of Chapter 4, we used JAGS instead of WinBUGS. One of the advantages is the fact that JAGS has cumulative distributions functions implemented, while in WinBUGS we have to write the expressions for $u = F(X)$ and $v = G(Y)$ for some distributions. This can be a problem when we have a Gamma distribution, for instance. Moreover, JAGS has quantile functions which in the case of Elliptical copulas are needed. Although JAGS has also a wider catalogue of distributions implemented, Burr is not one of them. In this case, one needs to note that a Burr distribution is a special case of a Pareto IV and, with careful adaptions, one is able to implement it when needed; see Appendix A. The Weibull and the Lognormal distributions have also different parameterisations which can be found on Appendix A.

It is worth mentioning the case of the Braga's station, addressed in Appendix B. During the period from May 9th of 2005 to April 30th of 2007, the values of daily maximum wind speed recorded followed a constant pattern. Since it is hardly possible to have exactly the same values, following the same sequence, of real wind speed data, we suspect it was due to a possible error of the meteorological station and, therefore, we have removed this period from the study.

Finally, some authors choose to fix the degrees of freedom of the Student *t*-Copula, instead of considering it as a random variable. For instance, it is not yet possible to obtain confidence intervals for this parameter if it is not fixed at first. Also, in the Bayesian approach, we have obtained very different estimates for this parameter and, in one of the cases, the length of the credible interval was huge.

Another limitation has to do with the 3-parameter Burr distribution. When fitting this distribution, we faced the problem of having to provide starting values for all parameters. Moreover, to the best of our knowledge, there is not a rule for choosing adequate values. Therefore, in some cases, the standard error of the first shape parameter as well as the range of its 95% confidence interval were very high. The same occurred when we performed the Bayesian inference. For instance, the MCMC simulation needed a high number of iterations in order to converge and the first shape parameter still showed some autocorrelation even after thinning.

A possible continuation of this work would be to develop simulation tools for other copulas, such as BB1, BB7, BB8, Tawn Type 1, Tawn Type 2 or Extreme-Value copulas. Therefore, we would work with at least more 40 possible copula families. Another possibility, in a Bayesian framework, could be to select the copula by using DIC, which we mentioned in Chapter 3 but did not addressed in this thesis.

References

[Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19(6): 716-723.

[Atique and Atooh-Okine, 2018] Atique, F. and Atooh-Okine, N. (2018). Copula Parameter Estimation Using Bayesian Inference for Pipe Data Analysis, *Canadian Journal of Civil Engineering*, 45(1): 61-70.

[Beirlant *et al*, 2004] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications.* John Wiley & Sons, Ltd, England.

[Berg, 2009] Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison, *The European Journal of Finance*, 15: 675-701.

[Boyd and Vandenberghe, 2004] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization.* Cambridge University Press, Cambridge, UK.

[Çelik and Yilmaz, 2008] Çelik, H. and Yilmaz, V. (2008). A Statistical Approach to Estimate the Wind Speed Distribution: The Case of Gelibou Region, *Dogus Üniversitesi Dergisi*, 9(1): 122-132.

[Cong and Brady, 2012] Cong, R.-G. and Brady, M. (2012). The Interdependence between Rainfall and Temperature: Copula Analyses, *The Scientific World Journal*, 2012: 1-11.

[Dana, 2007] Dana, L. K. (2007). Using Copulas to Model Dependence in Simulation Risk Assessment, *ASME Internations Mechanical Engineering Congress and Exposition, Proceedings*, 14.

[Delignette-Muller and Dutang, 2015] Delignette-Muller, M. and Dutang, C. (2015). Fitdistrplus: An R Package for Fitting Distributions, *Journal of Statistical Software*, 64: 1-34.

[Demarta and McNeil, 2005] Demarta, S. and McNeil, A. J. (2005). The *t* Copula and Related Copulas, *International Statistical Review*, 73(1): 111-129.

[Denwood, 2016] Denwood, M. (2016). runjags: An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS, *Journal of Statistical Software*, 71.

[Diaz, 2017] Díaz, M. G. Bayesian prediction of glacial discharge in Antartica using copulas (2017). *PhD Thesis*, Universidad Carlos III, Madrid, Spain.

[Dos Santos, 2011] Dos Santos, M. F. (2011). Aplicação de cópulas na modelação do número de sinistros de grupos de risco homogéneos no seguro automóvel. *Trabalho Académico*, ISEG, Universidade de Lisboa, Lisboa, Portugal.

[Dos Santos Silva and Lopes, 2008] Dos Santos Silva, R. and Lopes, H. (2008). Copula, marginal distributions and model selection: A Bayesian note. *Statistics and Computing*, 18: 313-320.

[Embrechts *et al*, 2003] Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling Dependence with Copulas and Applications to Risk Management, *Handbook of Heavy Tailed Distributions in Finance*. S.T. Rachev, Elsevier/North Holland, Amsterdam.

[Erntell, 2013] Erntell, F. On modeling insurance claims using copulas (2013). *Master's Thesis*, Lund University, Lund, Sweden.

[Gelman and Hill, 2006] Gelman, A. and Hill, J. (2006). *Data Analysis Using Regressian And Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, United Kingdom.

[Genest and Favre, 2007] Genest, C., Favre, A.-C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask, *Journal of Hydrologic Engineering*, 12(4): 347-368.

[Genest *et al*, 1995a] Genest, C., Ghoudi, K. and Rivest, L.-P. (1995a). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, 82(3): 543-552.

[Genest *et al*, 2009] Genest, C., Rémillard, B. and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study, *Insurance: Mathematics and Economics*, 44: 199-213.

[González-Estrada and Villaseñor, 2018] González-Estrada, E. and Villaseñor, J. A. (2018). An R package for testing goodness-of-fit: goft, *Journal of Statistical Computation and Simulation*, 88(4): 726-751.

[Harris and Cook, 2014] Harris, R. I. and Cook, N. J. (2014). The parent wind speed distribution: Why Weibull?. *Journal of Wind Engineering and Industrial Aerodynamics*, 131: 72-87.

[Hofert *et al*, 2018] Hofert, M., Kojadinovic, I, Mächler, M and Yan, J. (2018) *Elements of copula modeling with R.* Springer International Publishing.

[Huard *et al*, 2006] Huard, D., Évin, G. e Favre, A.-C. (2006). Bayesian copula selection. *Computational Statistics & Data Analysis*, 51: 809-822.

[Joe, 1997] Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Taylor & Francis Group, Florida, FL.

[Joe, 2005] Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based methods. *Journal of Multivariate Analysis*, 94(2): 401-419.

[Joe, 2014] Joe, H. (2014). *Dependence Modeling with Copulas.* Taylor & Francis Group, Florida, FL.

[Kostova *et al*, 2012] Kostova, S., Rumchev, K, Vlaev, T. and Popova, S. B. (2012). Using Copulas to Measure Association between Air Pollution and Respiratory Diseases, *Word Academy of Science, Engineering and Technology*, 6: 533-538.

[Lilliefors, 1967] Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *Journal of the American Statistical Association*, 62(318): 399-402.

[Littell *et al*, 1979] Littell, R. C., McClave, J. T. and Offen, W. W. (1979). Goodness-of-fit tests for the two parameter Weibull distribution, *Communications in Statistics - Simulation and Computation*, 8(3): 257-269.

[Marshall and Olkin, 1988] Marshall, A. W. and Olkin, I. (1988). Families of Multivariate Distributions, *Journal of the American Statistical Association*, 83(403): 834-841.

[Mert and Karakus, 2015] Mert, I. and Karakus, C. (2015). A statistical analysis of wind speed data using Burr, generalized gamma, and Weibull distributions in Antakya, Turkey, *Turkish Journal of Electrical Engineering and Computer Sciences*, 23: 1571-1586.

[Moore, 1986] Moore, D. S. (1986). Tests of Chi-Squared Type, *Goodness-of-Fit Techniques*. Marcel Dekker, New York: 63-95.

[Naveau *et al*, 2016] Naveau, P., Huser, R., Ribereau, P. and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, *Water Resources Research*, 52: 2753-2769.

[Nelsen, 2006] Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer Science+Business Media, New York, NY, 2nd Edition.

[Neto, 2015] Neto, J. (2015). Fitting Distributions.
`http://www.di.fc.ul.pt/~jpn/r/distributions/fitting.html`

[Novack-Gottshall and Wang, 2016] Novack-Gottshall, P. and Wang, S. (2016). Package “KScorrect” - Lilliefors-Corrected Kolmogorov-Smirnoff Goodness-of-Fit Tests.

[Okhrin *et al*, 2018] Okhrin, O., Trimborn, S., Zhang, S. and Zhou, Q. M. (2018). Goodness-of-Fit Tests for Copulae.

[Paulino *et al*, 2018] Paulino, C. D., Turkman, M. A. A., Murteira, B. and Silva, G.L. (2018). *Estatística Bayesiana.* Fundação Calouste Gulbenkian, Lisboa, 2ª Edição.

[Pobocikova *et al*, 2017] Pobocikova, I., Sedliackova, Z. and Michalková, M. (2017). Application of Four Probability Distributions for Wind Speed Modeling, *Procedia Engineering*, 192: 713-718.

[Rosenblatt, 1952] Rosenblatt, M. (1952). Remarks on a multivariate transformation, *The Annals of Mathematical Statistics*, 23: 470-472.

[Schwarz, 1978] Schwartz, G. E. (1978). Estimating the dimension of a model, *Annals of Statistics*, 6(2): 461-464.

[Schepsmeier *et al*, 2018] Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T. and Erhardt, T. (2018). VineCopula: Statistical Inference of Vine Copulas. R package version 2.1.8.

[Shemyakin and Kniazev, 2017] Shemyakin, A. and Kniazev, A. (2017). *Introduction to Bayesian Estimation and Copula Models of Dependence.* John Wiley & Sons, New York, NY.

[Shemyakin and Youn, 2006] Shemyakin, A. and Youn, H. (2006). Copula models of joint last survior analysis, *Applied Stochastic Models in Business and Industry*, 22: 211-224.

[Shepherd, 1978] Shepherd, D. G. (1978). Wind Power, *Advances in Energy Systems and Technology*, 1:1-124.

[Sklar, 1959] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs merges, *Publications de l'Institut de Statistique de l'Université de Paris*, 8: 229-231.

[Smith, 2011] Smith, M. S. (2011). Bayesian Approaches to Copula Modelling, *ERN: Bayesian Analysis (Topic)*.

[Spiegelhalter *et al*, 2002] Spiegelhalter, D. J., Best, N. G, Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society*, Series B, 64(4): 583-639.

[Stephens, 1986] Stephens, M. A. (1986). Tests Based on EDF Statistics, *Goodness-of-Fit Techniques*. Marcel Dekker, New York: 97-193.

[Tadikamalla, 1990] Tadikamalla, P. R. (1990). Kolmogorov-Smirnov Type Test-Statistics for the Gamma, Erlang-2 and the Inverse Gaussian Distributions When the Parameters are Unknown, *Communications in Statistics - Simulation and Computation*, 19(1): 305-314.

[The Beginner Programmer, 2015] The Beginner Programmer. (2015). How to fit a copula model in R. http://firsttimeprogrammer.blogspot.com/2015/02/how-to-fit-copula-model-in-r.html

[The Beginner Programmer, 2016] The Beginner Programmer. (2016). How to fit a copula model in R [heavily revised]. Part 1: basic tools. https://firsttimeprogrammer.blogspot.com/2016/03/how-to-fit-copula-model-in-r-heavily.html

[The Beginner Programmer, 2016] The Beginner Programmer. (2016). How to fit a copula model in R [heavily revised]. Part 2: fitting the copula. https://firsttimeprogrammer.blogspot.com/2016/03/how-to-fit-copula-model-in-r-heavily_29.html

[Wang and Wells, 2000] Wang, W and Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data, *Journal of the American Statistical Association*, 95(449): 62-72.

[Yan, 2007] Yan, J. (2007). Enjoy the Joy of Copulas: With a Package copula, *Journal of Statistical Software*, 21.

[Zscheischler *et al*, 2017] Zscheischler, J., Orth, R. and Seneviratne, S. (2017). Bivariate return periods of temperature and precipitation explain a large fraction of European crop yields, *Biogeosciences*, 14: 3309-3320.

A | Expressions

A.1 Lognormal Distribution

Let X be a random variable which follows a Lognormal distribution, that is $X \sim LN(\mu,\sigma)$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the mean and the standard deviation of $Y = \log(X)$, respectively. Thus, the *probability density function* is defined by

$$f(x \mid \mu,\sigma) = \frac{1}{x}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right\}, \quad x > 0. \tag{A.1}$$

In regard to the Bayesian inference, the precisian parameter, $\varepsilon = \frac{1}{\sigma^2}$, is considered instead of the standard deviation. Therefore, (A.1) is rewritten as

$$f(x \mid \mu,\varepsilon) = \frac{1}{x}\sqrt{\frac{\varepsilon}{2\pi}}\exp\left\{-\frac{\varepsilon(\log(x)-\mu)^2}{2}\right\}, \quad x > 0. \tag{A.2}$$

The *log-likelihood function* is given by

$$\log(L(\mu,\sigma \mid x)) = -\log(x\sigma\sqrt{2\pi}) - \frac{(\log(x)-\mu)^2}{2\sigma^2} \tag{A.3}$$

or equivalently by

$$\log(L(\mu,\varepsilon \mid x)) = \frac{1}{2}\log(\varepsilon) - \log(x\sqrt{2\pi}) - \frac{\varepsilon(\log(x)-\mu)^2}{2}. \tag{A.4}$$

$\mu \sim N(0,10^{-4})$, where 10^{-4} is the precision, and $\sigma \sim Gamma(10^{-3},10^{-3})$ are the non-informative priors distributions used for the Lognormal distribution.

A.2 Gamma Distribution

Let X be a random variable which follows a Gamma distribution, that is $X \sim Gamma(\alpha,\beta)$, where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the rate parameter. Thus, the *probability density function* is defined by

$$f(x \mid \alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp\{-\beta x\}, \quad x > 0, \tag{A.5}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x$ is the *Gamma function*.

The *log-likelihood function* is given by

$$\log(L(\alpha,\beta \mid x)) = \alpha\log(\beta) + (\alpha-1)\log(x) - \beta x - \log(\Gamma(\alpha)) \tag{A.6}$$

α, $\beta \sim Gamma(10^{-3},10^{-3})$ are the non-informative priors distributions used for the Gamma distribution.

A.3 Weibull Distribution

Let X be a random variable which follows a Weibull distribution, that is $X \sim \textit{Weibull}(\omega, \delta)$, where $\omega > 0$ is the shape parameter and $\delta > 0$ is the scale parameter. Thus, the *probability density function* is defined by

$$f(x \mid \omega, \delta) = \frac{\omega}{\delta}\left(\frac{x}{\delta}\right)^{\omega-1} \exp\left\{-\left(\frac{x}{\delta}\right)^{\omega}\right\}, \quad x \geq 0. \tag{A.7}$$

Setting $b = \delta^{\omega}$, the BUGS language considers the following parameterisation of the Weibull distribution:

$$f(x \mid \omega, b) = \omega b x^{\omega-1} \exp\{b x^{\omega}\}, \quad x \geq 0. \tag{A.8}$$

The *log-likelihood function* is given by

$$\log(L(\omega, \delta \mid x)) = \log(\omega) - \log(\delta) + (\omega - 1)[\log(x) - \log(\delta)] - \left(\frac{x}{\lambda}\right)^{\omega} \tag{A.9}$$

or equivalently by

$$\log(L(\omega, b \mid x)) = \log(\omega) + \log(b) + (\omega - 1)\log(x) - b x^{\omega}. \tag{A.10}$$

ω, δ (or b) $\sim \textit{Gamma}(10^{-3}, 10^{-3})$ are the non-informative priors distributions used for the Weibull distribution.

A.4 3-parameter Burr Distribution

Let X be a random variable which follows a 3-parameter Burr distribution, that is $X \sim \textit{Burr}(k, c, \lambda)$, where $k > 0$ and $c > 0$ are the shape parameters and $\lambda > 0$ is the rate parameter. Thus, the *probability density function* is defined by

$$f(x \mid k, c, \lambda) = k c \lambda (\lambda x)^{c-1} [1 + (\lambda x)^{c}]^{-(k+1)}, \quad x \geq 0. \tag{A.11}$$

The *log-likelihood function* is given by

$$\log(L(k, c, \lambda \mid x)) = \log(k) + \log(c) + \log(\lambda) + (c-1)[\log(x) + \log(\lambda)] - (k+1)\log\left[1 + (x\lambda)^{c}\right]. \tag{A.12}$$

k, c, $\lambda \sim \textit{Gamma}(10^{-3}, 10^{-3})$ are the non-informative priors distributions used for the 3-parameter Burr distribution.

If X follows a Pareto Type IV distribution, that is $X \sim \textit{PIV}(\alpha, \sigma, \mu, \gamma)$, where $\alpha > 0$ and $\gamma > 0$ are the shape parameters, $\sigma > 0$ is the scale parameter and $\mu \in \mathbb{R}$ is the location parameter. Then, the *probability density function* is defined by

$$f(x \mid \alpha, \sigma, \mu, \gamma) = \frac{\alpha \sigma^{-1/\gamma} (x - \mu)^{1/\gamma - 1} \left[1 + \left(\frac{\sigma}{x\mu}\right)^{-1/\gamma}\right]^{-(\alpha+1)}}{\gamma}, \quad x \geq \mu. \tag{A.13}$$

Setting $\alpha = k$, $\sigma = \frac{1}{\lambda}$, $\mu = 0$ and $\gamma = \frac{1}{c}$ in (A.13), we obtain (A.11). Therefore, the 3-parameter Burr distribution is a particular case of the Pareto Type IV distribution. This result will be important to apply the Bayesian inference.

A.5 Gaussian Copula

The log-likelihood function for the Gaussian copula with parameter ρ is defined by

$$\log(L(\rho \mid u, v)) = -\frac{1}{2}\log(1 - \rho^2) - \frac{\rho^2 s^2 + \rho^2 t^2 - 2\rho s t}{2(1 - \rho^2)}, \tag{A.14}$$

where $s = \Phi^{-1}(u)$ and $t = \Phi^{-1}(v)$. Since ρ is the Pearson's correlation coefficient, $\rho \sim U(-1, 1)$ is the non-informative prior distribution used.

A.6 Student t-Copula

The *log-likelihood function* for the Student t-Copula with correlation parameter ρ and η degrees of freedom is defined by

$$\begin{aligned}\log(L(\rho,\eta \mid u,v)) =& \log\left(\Gamma\left(\frac{\eta+2}{2}\right)\right)+\log\left(\Gamma\left(\frac{\eta}{2}\right)\right)-\frac{1}{2}\log(1-\rho^2)-2\log\left(\Gamma\left(\frac{\eta+1}{2}\right)\right)\\ &+\frac{\eta+1}{2}\left[\log\left(1+\frac{s^2}{\eta}\right)+\log\left(1+\frac{t^2}{\eta}\right)\right]-\frac{\eta+2}{2}\log\left[1+\frac{s^2+t^2-2\rho st}{\eta(1-\rho^2)}\right], \qquad \text{(A.15)}\end{aligned}$$

where $s=T_\eta^{-1}(u)$ and $t=T_\eta^{-1}(v)$. Since ρ is the Pearson's correlation coefficient, $\rho \sim U(-1,1)$ is the non-informative prior distribution used. BUGS language restricts the degrees of freedom parameters to be at least 2. Therefore, setting $\nu=\frac{1}{\eta}$, we can use $\nu \sim U\left(0,\frac{1}{2}\right)$ as a non-informative prior distribution for "η"; see [Gelman and Hill, 2006].

A.7 Survival Clayton Copula

The *log-likelihood function* for the Survival Clayton copula with association parameter α is defined by

$$\begin{aligned}\log(L(\alpha \mid u,v)) =& \log\left(1+\frac{1}{\alpha}\right)+\log(\alpha)+(\alpha+1)[\log(1-u)-\log(1-v)]\\ &-\left(2+\frac{1}{\alpha}\right)\log[(1-u)^{-\alpha}+(1-v)^{-\alpha}-1]. \qquad \text{(A.16)}\end{aligned}$$

Recalling the relationship between the Clayton association parameter and Kendall's tau, that is, $\alpha=\frac{2\tau}{1-\tau}$, which is valid for its survival version, we note that τ tends to 1 as α increases to ∞. Therefore, $\alpha \sim U(0,200)$ is the non-informative prior distribution used.

A.8 Frank Copula

The *log-likelihood function* for the Frank copula with association parameter α is defined by

$$\log(L(\alpha \mid u,v)) = \log(\alpha)+\log(1-e^{-\alpha})-\alpha(u+v)-2\log[1-e^{-\alpha}-(1-e^{-\alpha u})(1-e^{-\alpha v})]. \qquad \text{(A.17)}$$

τ approaches 1 as α increases to ∞, and thus $\alpha \sim U(0,400)$ is used as a non-informative prior to the copula parameter.

A.9 Gumbel-Hougaard Copula

The *log-likelihood function* for the Gumbel-Hougaard copula with association parameter α is defined by

$$\log(L(\alpha \mid u,v)) = -\log(uv)+(\alpha-1)\log(\log(u)\log(v))+\log\left[w^{\frac{2}{\alpha}-2}+(\alpha-1)w^{\frac{1}{\alpha}-2}\right]-w^{\frac{1}{\alpha}}, \qquad \text{(A.18)}$$

where $w=(-\log(u))^\alpha+(-\log(v))^\alpha$. For this copula, $\tau=\frac{1}{1-\alpha}$. Therefore, if we set $\theta=\frac{1}{\alpha}$, τ tends to 1 if θ approaches 0. Thus, $\theta \sim Beta\left(\frac{1}{2},\frac{1}{2}\right)$ was used as a non-informative prior distribution for the copula parameter.

A.10 Joe Copula

The *Joe Copula* is an Archimedean copula with generator $\varphi_\alpha(t) = -\log[1-(1-t)^\alpha]$ and pseudo-inverse $\varphi_\alpha^{[-1]}(s) = 1-(1-e^{-s})^{\frac{1}{\alpha}}$. The Joe copula is given by

$$C_\alpha(u,v) = 1-[(1-u)^\alpha + (1-v)^\alpha - (1-u)^\alpha(1-v)^\alpha]^{\frac{1}{\alpha}}, \quad \alpha \in [1,\infty), \tag{A.19}$$

and its density is given by

$$c_\alpha(u,v) = [w^\alpha + z^\alpha - wz^\alpha]^{\frac{1}{\alpha}-2}\, wz^{\alpha-1}\,[\alpha - 1 + w^\alpha + z^\alpha - wz^\alpha], \quad \alpha \in [1,\infty), \tag{A.20}$$

where $w = 1-u$ and $z = 1-v$. It is characterised by having upper tail dependence. Moreover, $\lambda_U = 2-2^{\frac{1}{\alpha}}$.

Kendall's tau can be determined by

$$\tau = 1 + \frac{2}{2-\alpha}\left[\digamma(2) - \digamma\left(\frac{2}{\alpha}+1\right)\right], \tag{A.21}$$

where $\digamma(x) = \frac{d}{\mathrm{d}x}\log(\Gamma(x))$ is the digamma function; see [Joe, 2014].

BB1, Survival BB1, BB7, Survival BB7, BB8, Survival BB8 and other families of copulas are presented in detail in [Joe, 2014].

B | Results

B.1 The Case of Braga

In the process of analysing all data, we came across with a problem concerning Braga's meteorological station. From May 9th of 2005 to April 30th of 2007 the values of daily maximum wind speed registered seemed to follow a pattern, as we can see in Fig B.1. We might expect to happen a similar problem in the data produced by the simulator, due to some error of the model or of the machine. However, it is impossible that the daily maximum wind in the region of Braga took exactly the same values, for almost 2 years. Therefore, beyond the data preprocessing mentioned in Section 4.2, we have also removed the observations of this period. For this reason, the sample size of each season is low; see Tab. B.1

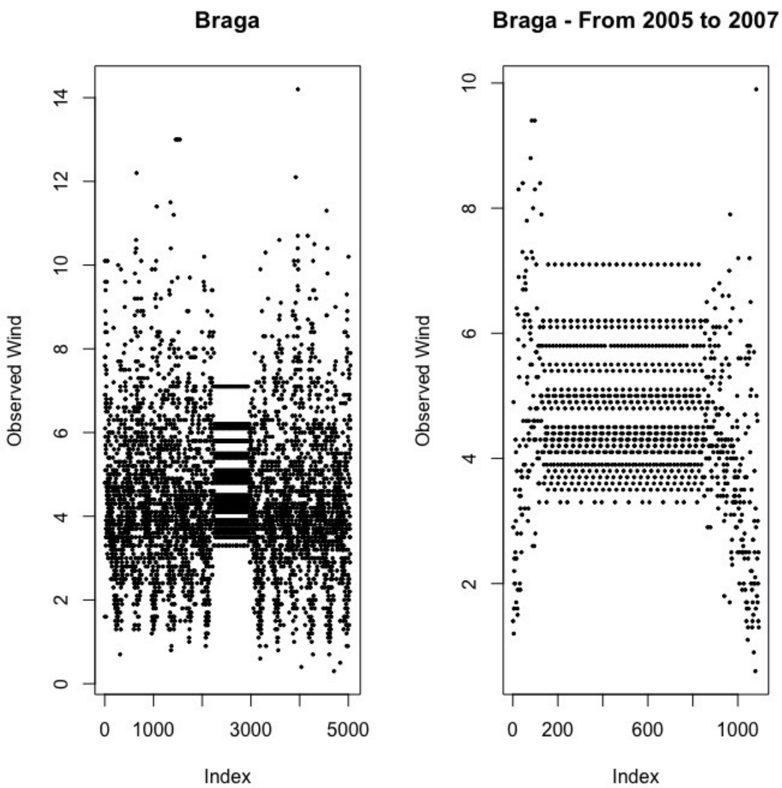


Fig. B.1: Observed wind speed in Braga's station

Tab. B.1: Summary of the values of X and Y in Braga

	Autumn							**Winter**						
	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
X	109	1.1	2.7	3.5	3.96	5.1	10.70	40	1	2.875	4.05	4.42	6.025	9.2
Y		1.401	4.007	5.04	5.381	6.456	13.036		2.009	3.606	5.64	5.784	7.202	13.124
	Spring							**Summer**						
	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
X	79	2.1	4	4.7	4.804	5.5	8	108	1.7	3.8	5.3	6.13	7.6	18.5
Y		2.848	6.064	7.076	6.938	7.863	9.703		1.7	3.4	3.9	4.005	4.3	7

In Tab. B.2, the goodness-of-fit tests for Braga's meteorological station are presented, and in Tab. B.3 the chosen marginal distributions, their maximum likelihood estimates and their 95% confidence intervals are shown. The marginal distributions fitted to the variables are in Fig. B.2.

Tab. B.2: Goodness-of-fit tests' results for the marginal distributions fitted to Braga's meteorological station. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

Autumn															
$F(x)$	p-val. χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-val. χ^2	KS	AD	CvM	S-W	AIC	R^2
LN	0.01	not rej.	–	–	rej.	443.7416	0.9791	LN	0.0036	not rej.	–	–	rej.	482.1502	0.9768
G	0.0112	not rej.	not rej.	not rej.	–	**441.897**	0.9869	G	0.0234	not rej.	not rej.	not rej.	–	**479.32**	0.99
W	0.0185	rej.	rej.	rej.	–	447.1812	0.9623	W	0.0225	rej.	not rej.	rej.	–	485.452	0.9699
B	0.0107	–	–	–	–	445.7701	0.9841	B	0.0325	–	–	–	–	481.8066	0.9931
Winter															
$F(x)$	p-val. χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-val. χ^2	KS	AD	CvM	S-W	AIC	R^2
LN	0.2682	not rej.	–	–	not rej.	177.761	0.9685	LN	0.1213	not rej.	–	–	not rej.	186.8161	0.9772
G	0.5534	not rej.	not rej.	not rej.	–	175.1546	0.9783	G	0.1516	not rej.	not rej.	not rej.	–	**185.9554**	0.9765
W	0.7705	not rej.	not rej.	not rej.	–	**174.377**	0.9886	W	0.1161	not rej.	not rej.	not rej.	–	187.5711	0.9707
B	0.2681	–	–	–	–	177.9475	0.9782	B	0.077	–	–	–	–	188.6833	0.9771
Spring															
$F(x)$	p-val. χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-val. χ^2	KS	AD	CvM	S-W	AIC	R^2
LN	0.7846	not rej.	–	–	not rej.	264.2931	0.9724	LN	0.021	rej.	–	–	rej.	294.2077	0.9118
G	0.7682	not rej.	not rej.	not rej.	–	**262.0509**	0.99	G	0.595	rej.	rej.	rej.	–	287.8696	0.9594
W	0.186	not rej.	rej.	rej.	–	266.1558	0.9757	W	0.4331	not rej.	not rej.	not rej.	–	**275.1514**	0.9866
B	0.761	–	–	–	–	263.5774	0.9896	B	0.3161	–	–	–	–	277.2699	0.9942
Summer															
$F(x)$	p-val. χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-val. χ^2	KS	AD	CvM	S-W	AIC	R^2
LN	0.1631	rej.	–	–	rej.	285.6265	0.953	LN	0.2469	rej.	–	–	rej.	367.0257	0.9195
G	0.1119	rej.	rej.	rej.	–	285.8731	0.9347	G	0.4934	not rej.	rej.	rej.	–	360.1219	0.9706
W	0.0001	rej.	rej.	rej.	–	305.0802	0.927	W	0.7473	not rej.	not rej.	not rej.	–	353.4842	0.9779
B	0.3709	–	–	–	–	**280.2705**	0.9728	B	0.9271	–	–	–	–	**350.5758**	0.961

Tab. B.3: Fitted distributions to the observed, X, and simulated, Y, winds of Braga's station. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn						**Winter**					
X	$\hat{\theta}_{MLE}$	CI (95%)	Y	$\hat{\theta}_{MLE}$	CI (95%)	X	$\hat{\theta}_{MLE}$	CI (95%)	Y	$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\alpha}$	4.0449	(3.2068,5.4534)	$\hat{k}$	1.9707	(0.9054,45.1084)	$\hat{\omega}$	2.2222	(1.8001,2.9495)	$\hat{\alpha}$	5.9991	(5.1258,7.2966)
$\hat{\beta}$	1.0216	(0.7984,1.3715)	$\hat{c}$	3.3691	(2.6011,4.5909)	$\hat{\delta}$	5.0047	(4.2692,5.7374)	$\hat{\beta}$	7.4853	(7.1896,7.7808)
–	–	–	$\hat{\lambda}$	0.1525	(0.0435,0.2112)	–	–	–	–	–	–
Spring						**Summer**					
X	$\hat{\theta}_{MLE}$	CI (95%)	Y	$\hat{\theta}_{MLE}$	CI (95%)	X	$\hat{\theta}_{MLE}$	CI (95%)	Y	$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\alpha}$	14.3362	(11.6957,20.4242)	$\hat{\omega}$	19.1999	(16.2291,23.0993)	$\hat{k}$	0.8813	(0.4759,2.2363)	$\hat{k}$	3.4617	(1.2778,163.4954)
$\hat{\beta}$	2.9844	(2.2889,4.2843)	$\hat{\delta}$	2.4445	(2.0599,2.9435)	$\hat{c}$	8.6377	(6.4959,12.1571)	$\hat{c}$	7.6986	(6.205,10.2271)
–	–	–	–	–	–	$\hat{\lambda}$	0.2622	(0.2174,0.2895)	$\hat{\lambda}$	0.1194	(0.0646,0.14221)

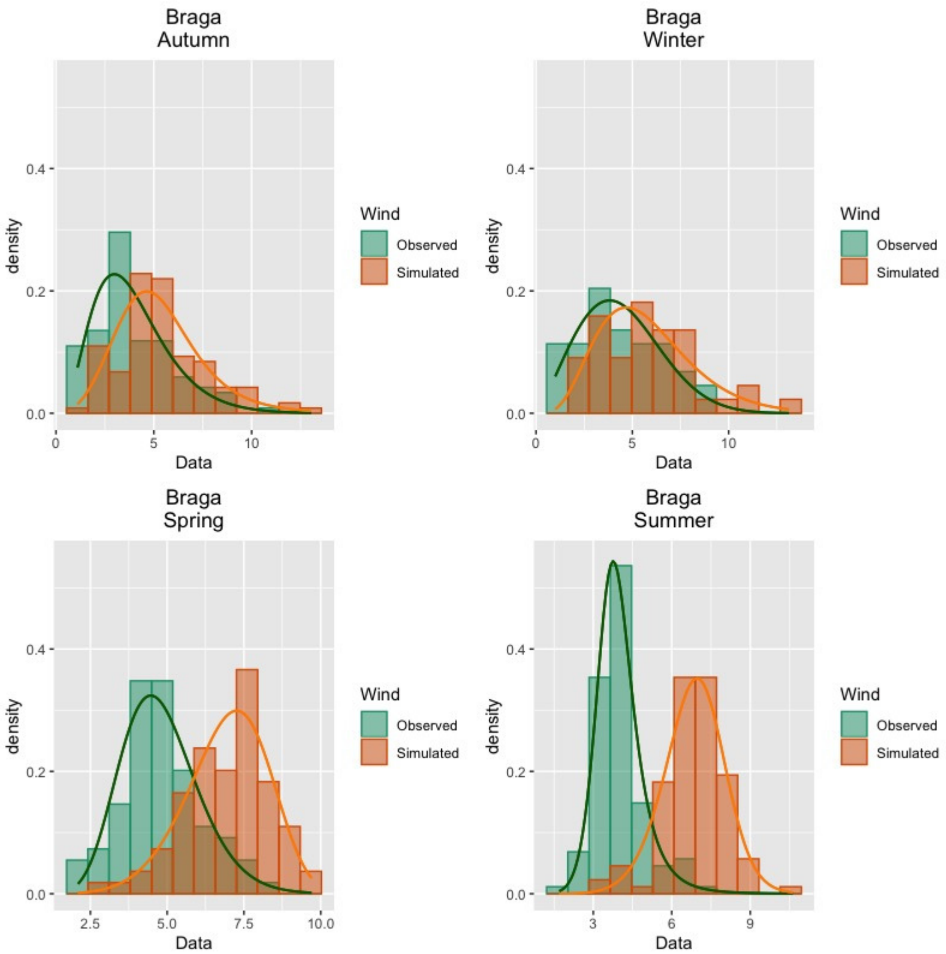


Fig. B.2: Marginal distributions fitted to the 4 seasons of Braga

The AIC was used to chose the best copula to fit the dependence of the wind speeds of each season and the copula parameters were estimated by the maximum pseudo-likelihood method; see Tab. B.4. All four copulas have tail dependence and none was rejected by any goodness-of-fit test. For instance, just the copula fitted to Spring has lower tail dependence as well. We can see that, in Autumn and Winter, the observed and the simulated wind speeds are highly correlated for high values; see Fig. B.3.

Tab. B.4: Copulas selected according to the AIC to Braga's station. $\hat{\theta}_1$ and $\hat{\theta}_2$ are the copula parameters estimates obtained by MPLE.

Season	Copula	$\hat{\theta}_{MPLE}$				Tail Dependence[(1)]		p-values		
		$\hat{\theta}_1$	CI (95%)	$\hat{\theta}_2$	CI (95%)	$\hat{\lambda}_L$	$\hat{\lambda}_U$	$S_n^{(B)}$	S_n	$S_n^{(K)}$
Autumn	C_α^{sc}	2.4986	(1.8111,3.1861)	–	–	0	0.7577	0.523	0.24	0.95
Winter	C_α^{j}	4.6704	(2.197,7.1439)	–	–	0	0.84	– (**)	0.0814	0.78
Spring	$C_{\rho\eta}^{t}$	0.4742	(0.2616,0.6868)	2.0701	(*)	0.3706	0.3706	1	0.8317	0.22
Summer	C_α^{sc}	1.0407	(0.6202,1.4612)	–	–	0	0.5137	0.387	0.26	0.63

(*) At present the asymptotic variance cannot be fully estimated if η is not fixed, thus it is not possible to provide a confidence interval; see the package *copula* manual. (**) $S_n^{(B)}$ is not implemented for the Joe Copula; see the package *gofCopula* manual. [(1)] the 0's are theoretical.

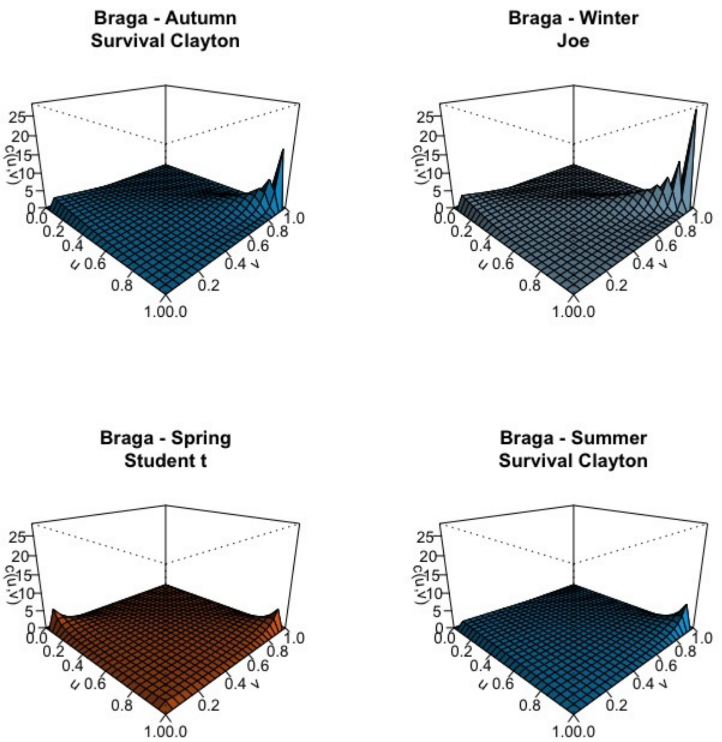


Fig. B.3: Probability density function of the fitted copulas to Braga.

B.2 Figures of the Remaining 8 Selected Meteorological Stations

We present now the plots of the ACF, the marginal distribution functions and the probability density of the copulas for the remaining selected meteorological stations: Aveiro, Bragança, Coruche, Estremoz, Lisboa S1, Monção, Sines and Vila do Bispo.

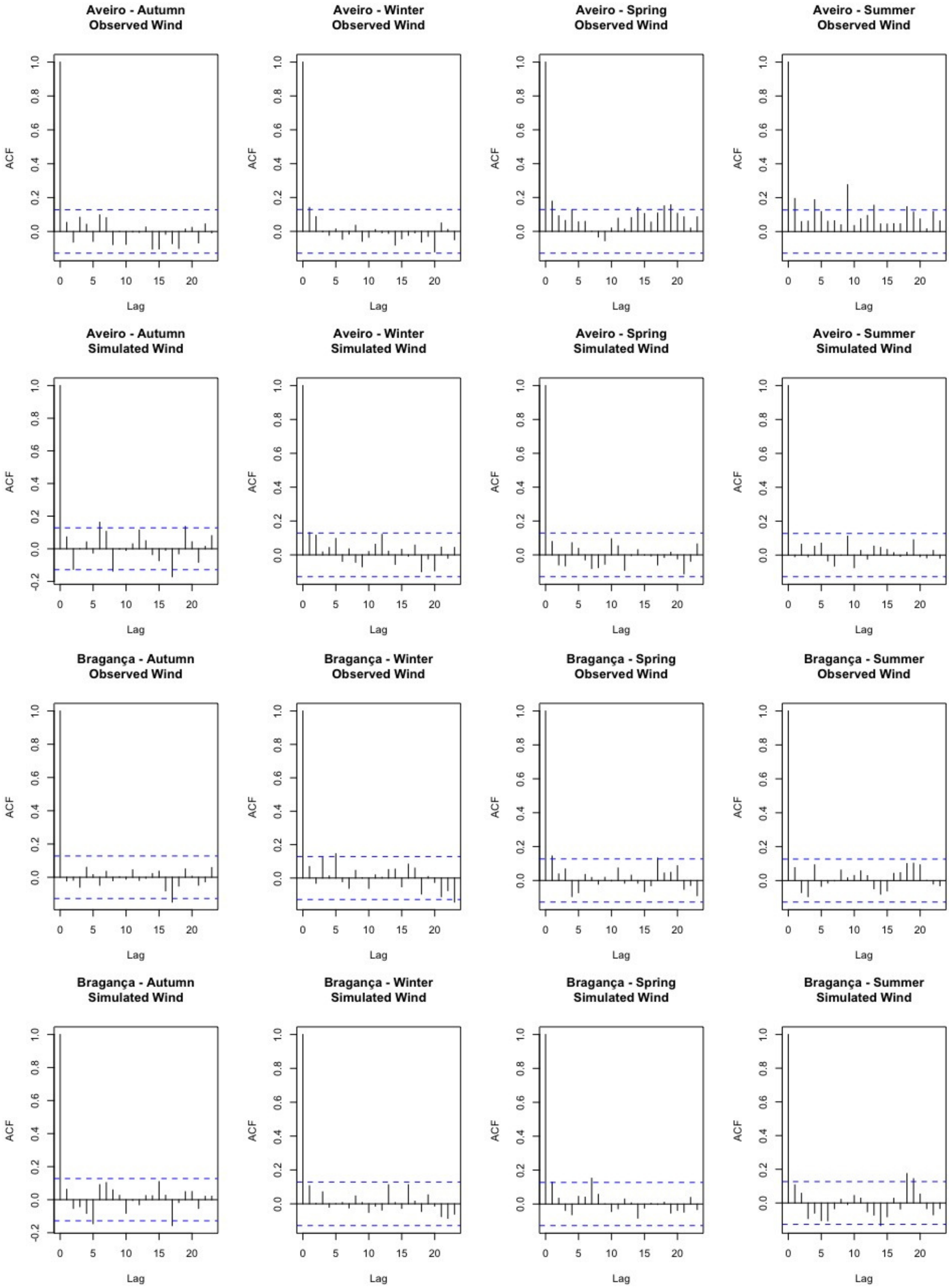


Fig. B.4: ACF of the observed and simulated wind speed data for each season of the meteorological stations of Aveiro and Bragança.

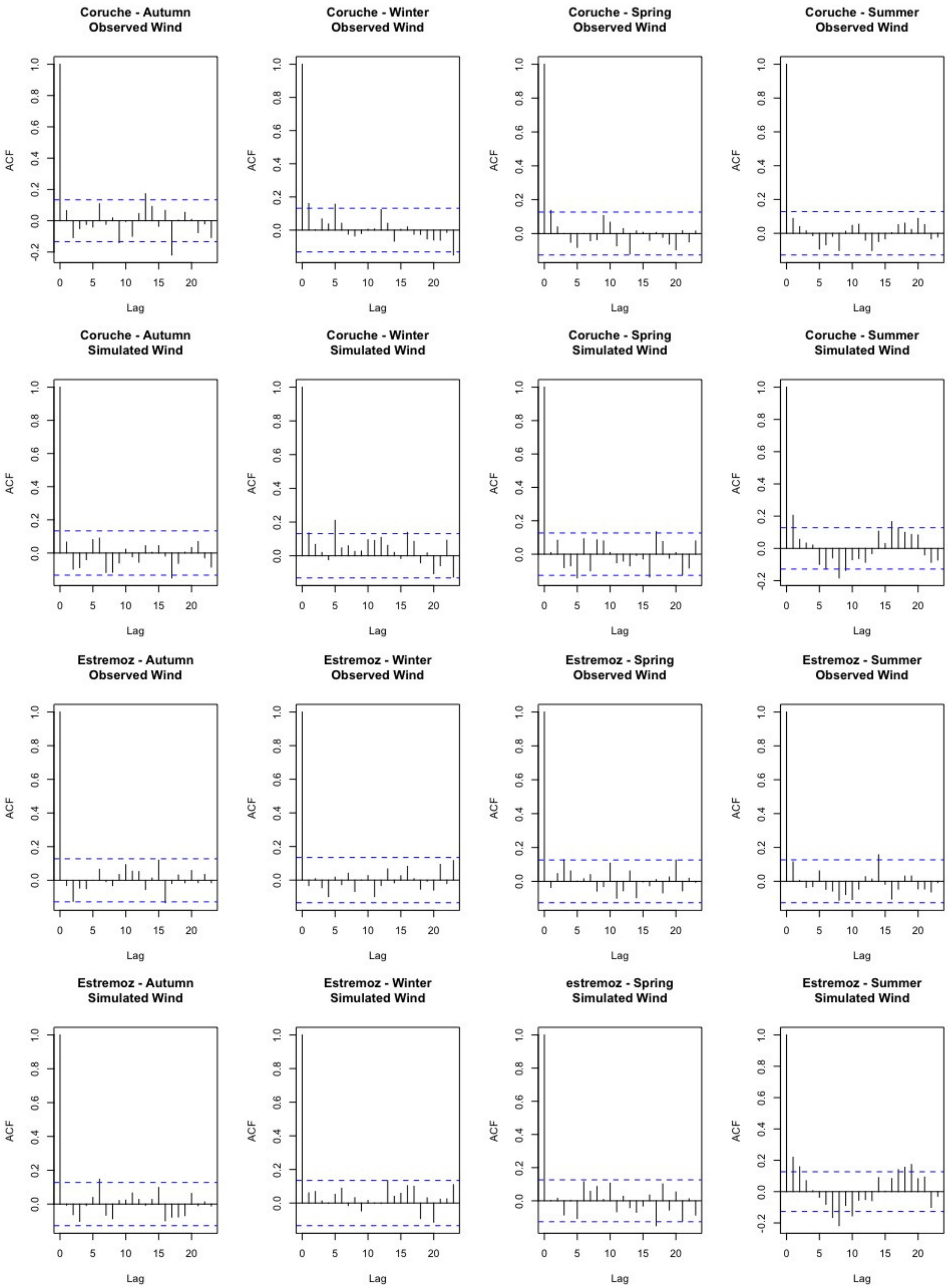


Fig. B.5: ACF of the observed and simulated wind speed data for each season of the meteorological stations of Coruche and Estremoz.

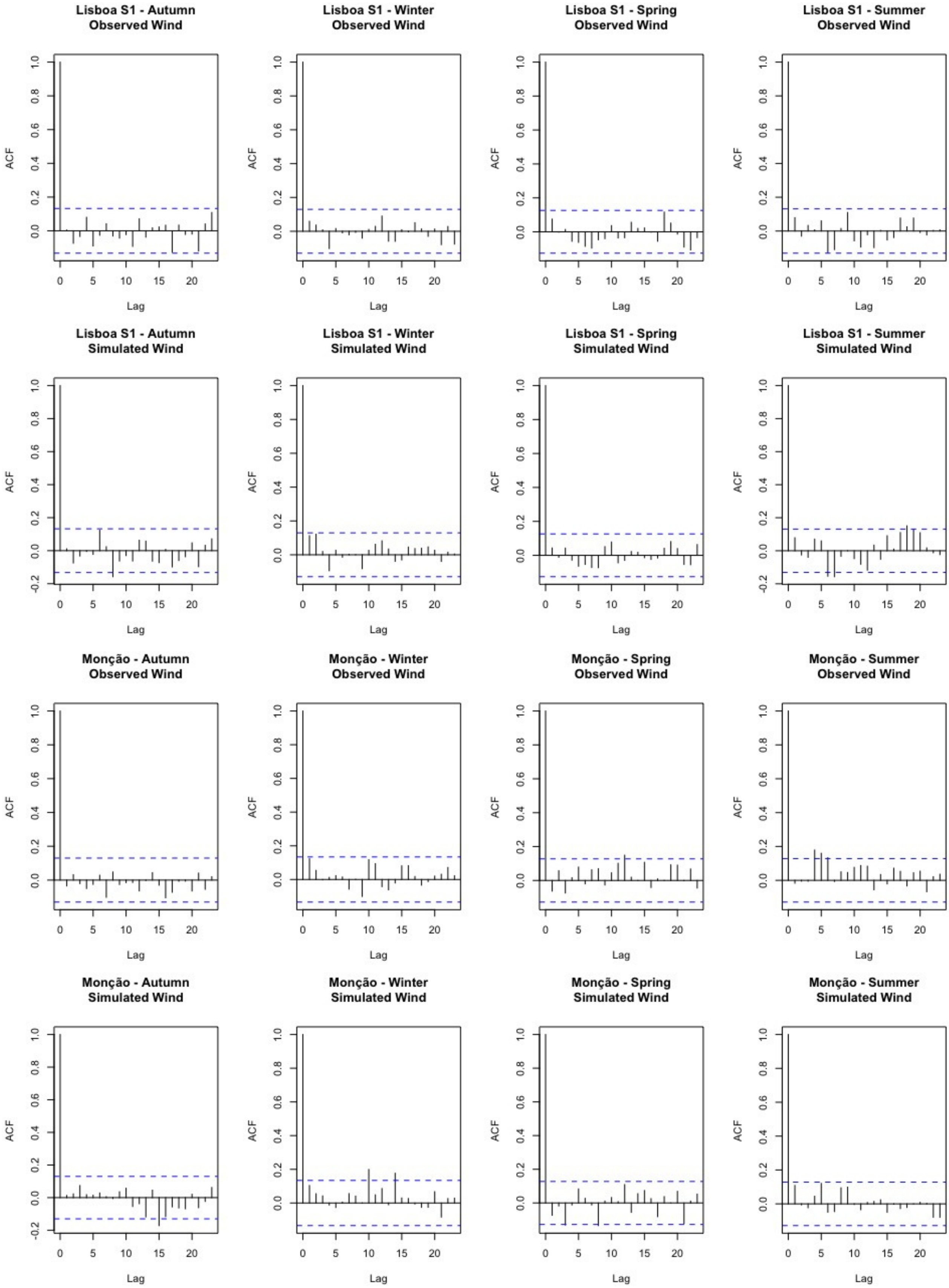


Fig. B.6: ACF of the observed and simulated wind speed data for each season of the meteorological stations of Lisboa S1 and Monção.

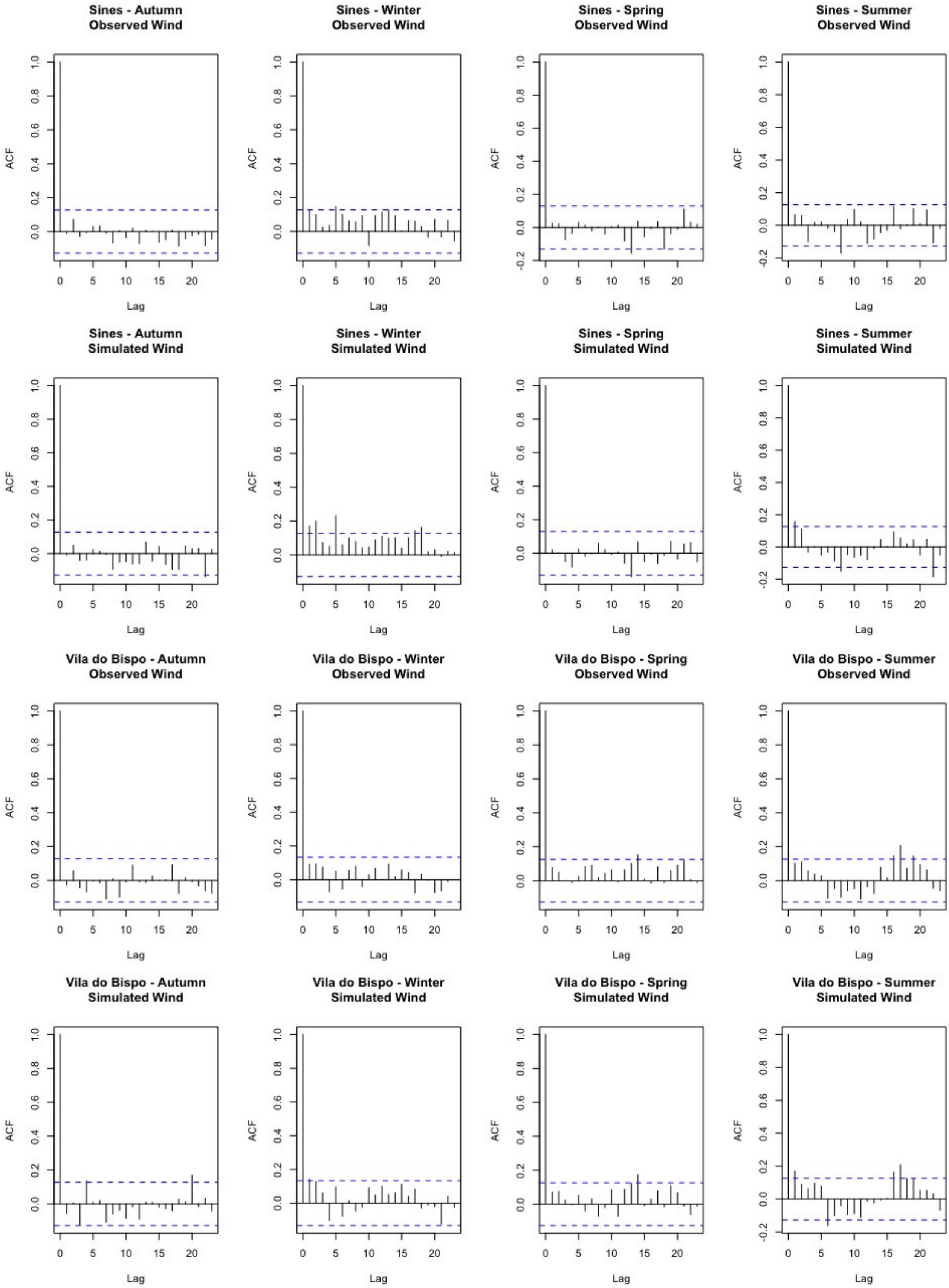


Fig. B.7: ACF of the observed and simulated wind speed data for each season of the meteorological stations of Sines and Vila do Bispo.

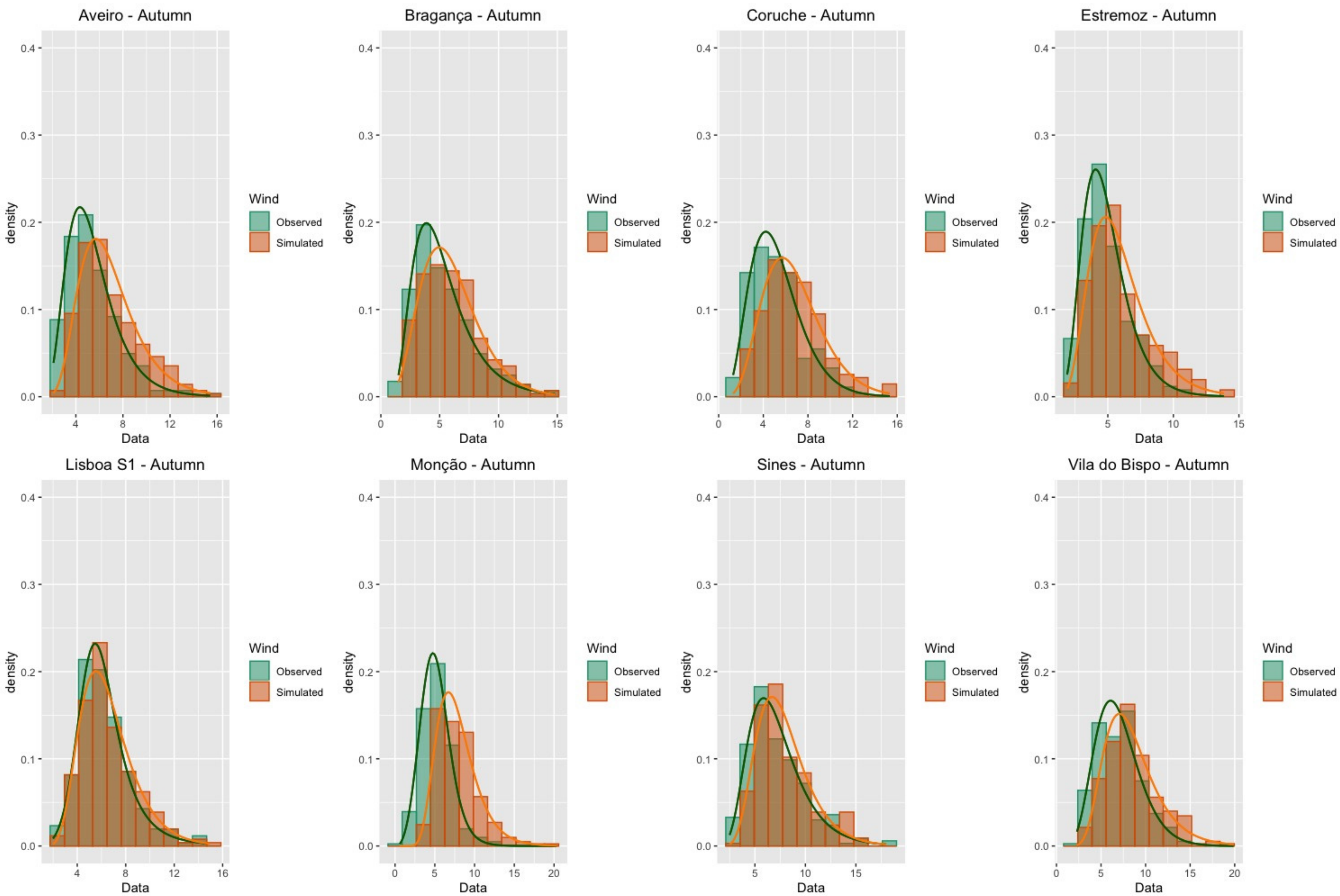


Fig. B.8: Marginal distributions fitted to the remaining 8 stations in Autumn.

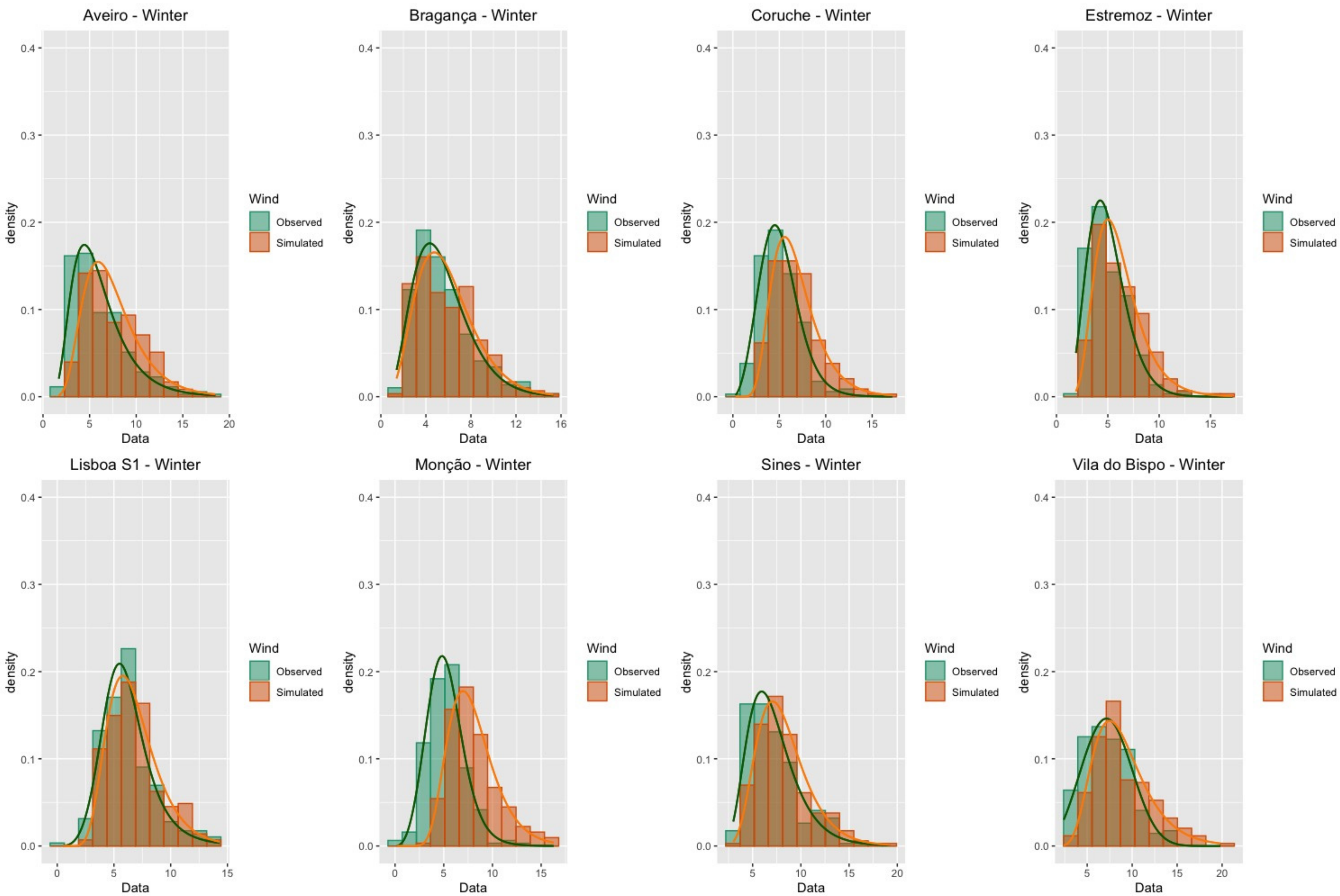


Fig. B.9: Marginal distributions fitted to the remaining 8 stations in Winter.

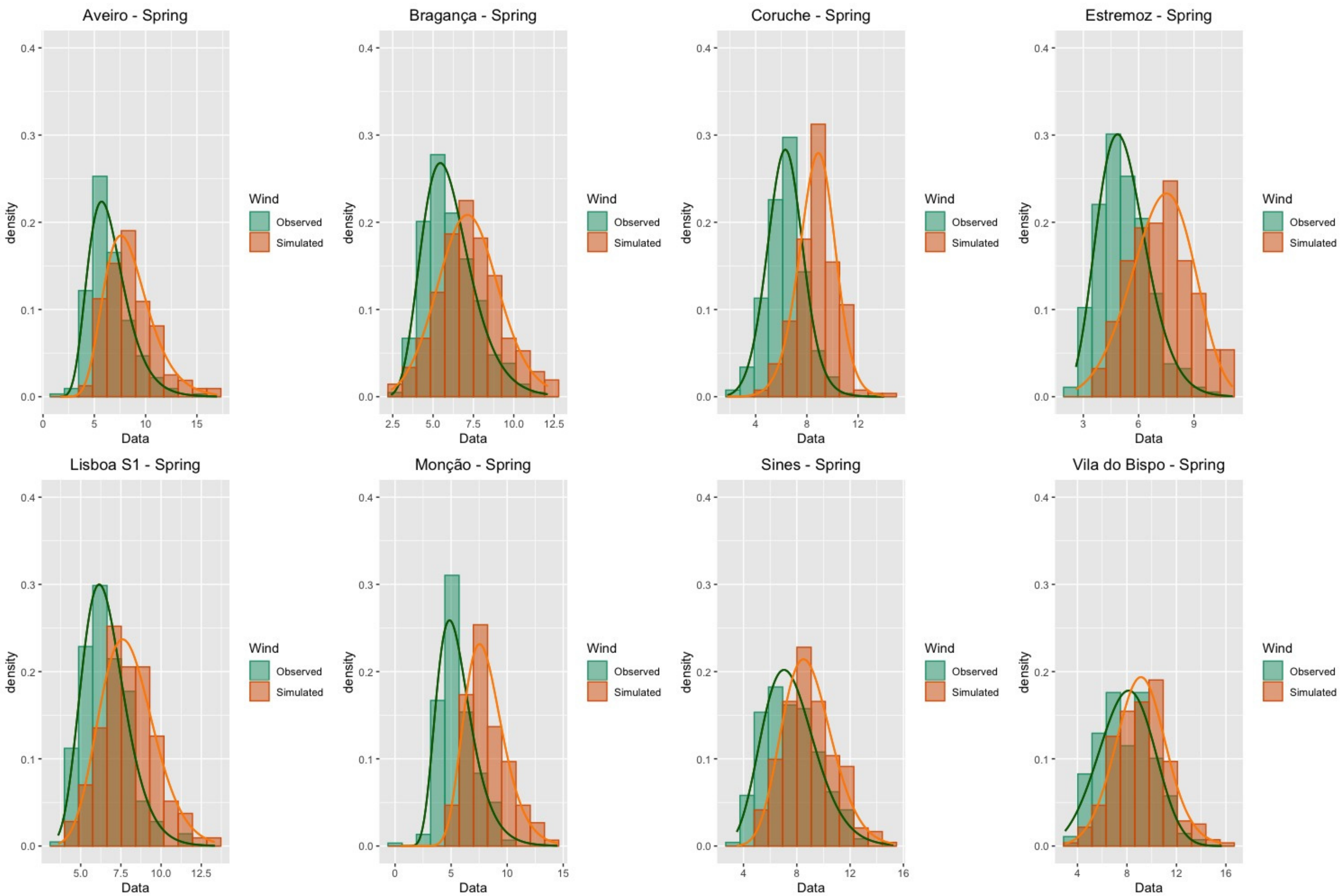


Fig. B.10: Marginal distributions fitted to the remaining 8 stations in Spring.

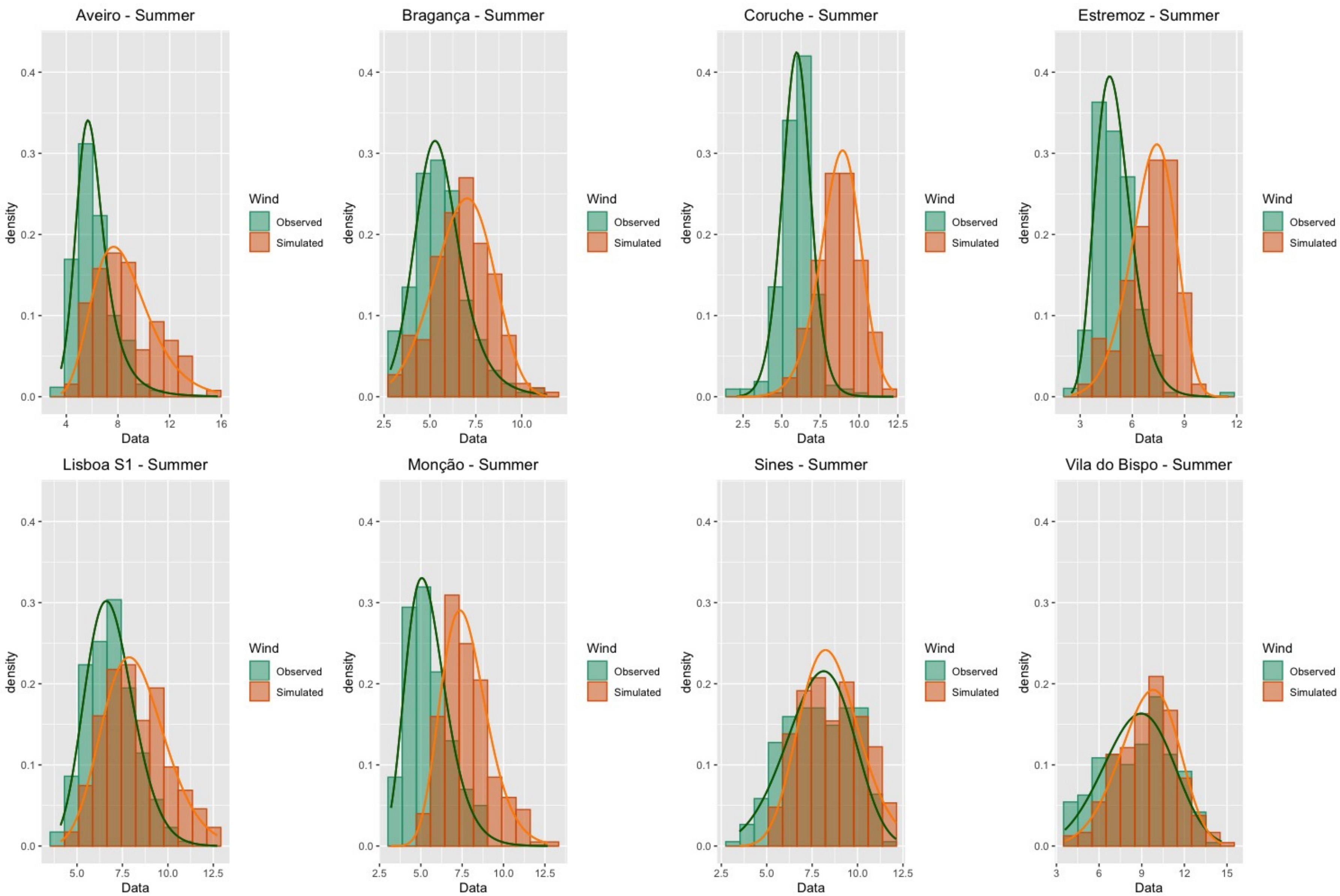


Fig. B.11: Marginal distributions fitted to the remaining 8 stations in Summer.

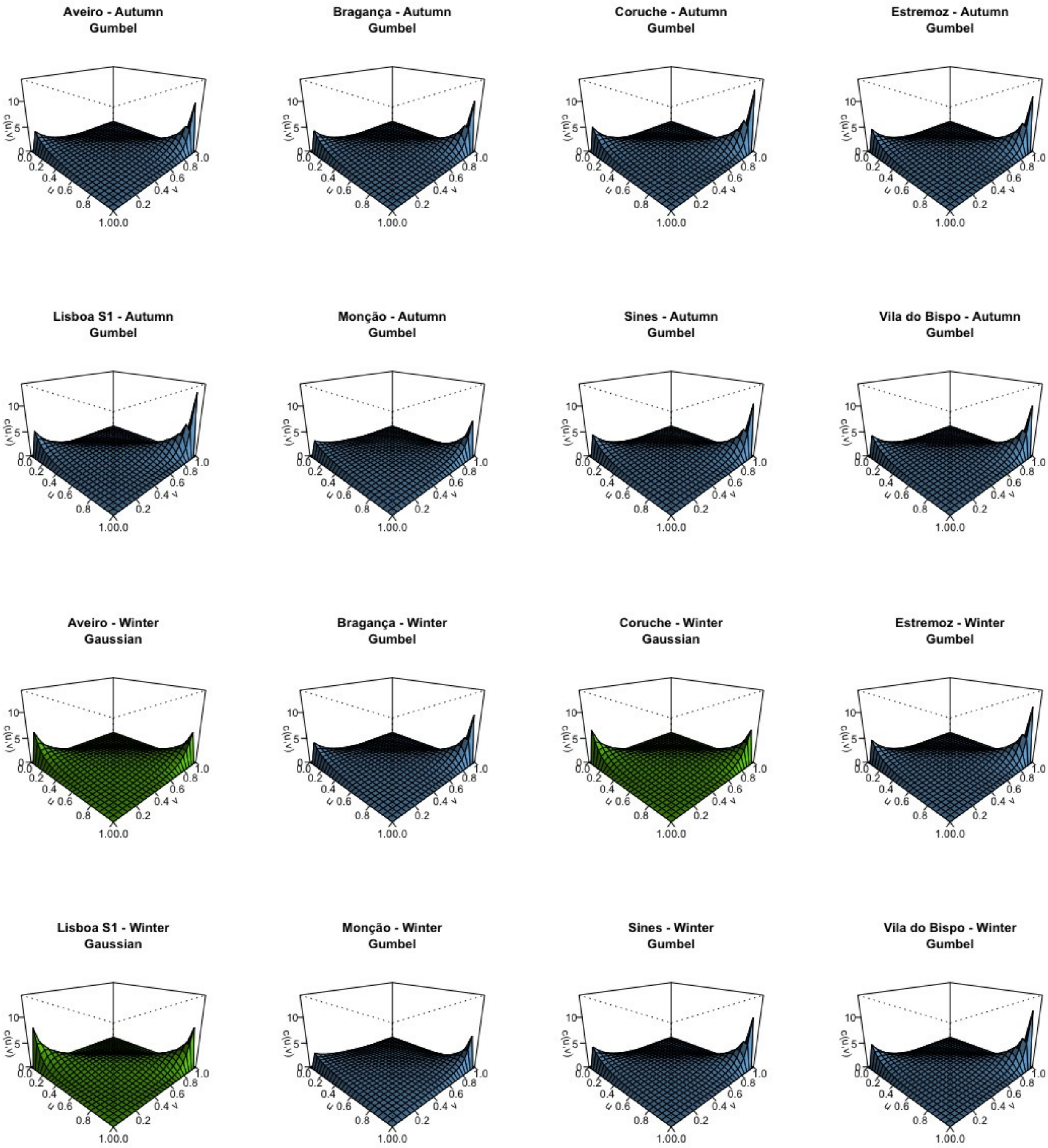


Fig. B.12: Probability density function of the fitted copulas to the remaining 8 stations in Autumn and Winter.

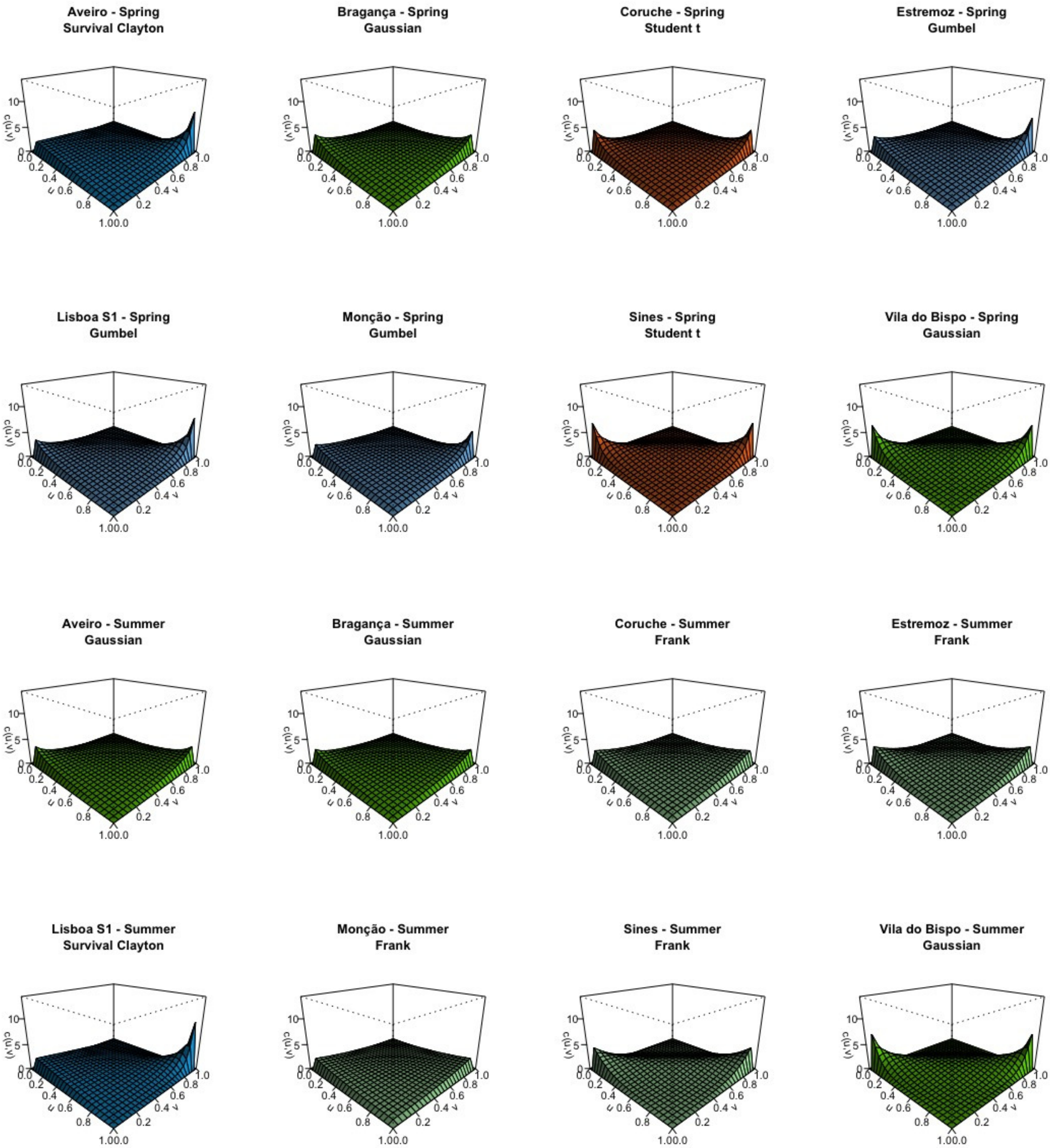


Fig. B.13: Probability density function of the fitted copulas to the remaining 8 stations in Spring and Summer.

B.3 Results for the Remaining 30 Meteorological Stations

In this section we summarise the results for the remaining 30 meteorological stations. We present the results of the goodness-of-fit tests for the marginal distributions, the fitted marginal distributions and the fitted copulas to each season of the remaining 30 meteorological stations of our work.

Tab. B.5: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Alcácer do Sal, Alcobaça, Barreiro, Beja, Cascais, Castro Verde, Chaves and Coimbra, in Autumn. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Autumn															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Alcácer do Sal	LN	0.1197	rej.	–	–	rej.	1006.3937	0.9857	LN	0.0316	not rej.	–	–	rej.	1052.5015	0.9859
	G	*0.3028*	*not rej.*	*not rej.*	*not rej.*	–	***999.9199***	*0.9898*	G	*0.0688*	*not rej.*	*not rej.*	*not rej.*	–	***1047.8802***	*0.9773*
	W	0.0946	rej.	rej.	rej.	–	1010.0659	0.9775	W	0.0097	not rej.	rej.	rej.	–	1057.1985	0.9667
	B	0.2624	–	–	–	–	1005.1638	0.9866	B	0.0463	–	–	–	–	1055.4683	0.9744
Alcobaça	LN	0.7706	not rej.	–	–	rej.	827.1373	0.9716	LN	0.0517	not rej.	–	–	rej.	**1133.8963**	0.9921
	G	0.3246	rej.	rej.	rej.	–	828.6485	0.9675	G	*0.0555*	*not rej.*	*rej.*	*rej.*	–	1135.629	*0.993*
	W	0.0001	rej.	rej.	rej.	–	863.123	0.9384	W	0.0012	rej.	rej.	rej.	–	1157.0379	0.9411
	B	*0.9787*	–	–	–	–	***819.0559***	*0.9884*	B	0.0126	–	–	–	–	1145.9156	0.9903
Barreiro	LN	*0.122*	*not rej.*	–	–	*not rej.*	***918.267***	*0.9929*	LN	*0.6103*	*not rej.*	–	–	*not rej.*	***1061.8273***	*0.9976*
	G	0.1155	not rej.	not rej.	not rej.	–	921.4762	0.9728	G	0.4623	not rej.	not rej.	not rej.	–	1065.1201	0.9823
	W	0.0002	rej.	rej.	rej.	–	953.8945	0.9531	W	0.0058	rej.	rej.	rej.	–	1092.726	0.9545
	B	0.1448	–	–	–	–	919.3974	0.9946	B	0.4463	–	–	–	–	1068.2714	0.9942
Beja	LN	*0.2259*	*not rej.*	–	–	*rej.*	***1051.6874***	*0.9906*	LN	*0.0972*	*not rej.*	–	–	*rej.*	***1082.904***	*0.9865*
	G	0.1951	not rej.	rej.	rej.	–	1059.3859	0.9932	G	0.0194	rej.	rej.	rej.	–	1090.9517	0.9909
	W	0.0089	rej.	rej.	rej.	–	1085.6767	0.9181	W	0.0001	rej.	rej.	rej.	–	1114.9506	0.9129
	B	0.0927	–	–	–	–	1066.8149	0.9767	B	0.0061	–	–	–	–	1100.2418	0.9595
Cascais	LN	0.0306	not rej.	–	–	not rej.	1029.7422	0.9927	LN	0.0085	not rej.	–	–	rej.	1064.2777	0.9844
	G	*0.0295*	*not rej.*	*not rej.*	*not rej.*	–	***1028.0083***	*0.9917*	G	*0.0185*	*not rej.*	*not rej.*	*not rej.*	–	***1059.4343***	*0.9836*
	W	0.0006	rej.	rej.	rej.	–	1043.667	0.9676	W	0.0033	rej.	rej.	rej.	–	1065.1914	0.9664
	B	0.009	–	–	–	–	1034.5658	0.9935	B	0.0043	–	–	–	–	1066.0694	0.9914
Castro Verde	LN	0.5305	not rej.	–	–	not rej.	**991.736**	0.9966	LN	*0.3941*	*not rej.*	–	–	*not rej.*	***1078.5259***	*0.9952*
	G	*0.4254*	*not rej.*	*not rej.*	*not rej.*	–	***994.2064***	*0.9923*	G	0.1184	not rej.	rej.	not rej.	–	1084.0849	0.9925
	W	0.0132	rej.	rej.	rej.	–	1017.7386	0.9504	W	0.0002	rej.	rej.	rej.	–	1110.8738	0.9335
	B	0.3595	–	–	–	–	1001.7221	0.9915	B	0.1286	–	–	–	–	1091.6694	0.9841
Chaves	LN	0.0249	rej.	–	–	rej.	858.2173	0.9856	LN	0.3003	not rej.	–	–	rej.	1046.8155	0.9897
	G	*0.1677*	*not rej.*	*rej.*	*not rej.*	–	***855.8195***	*0.9792*	G	*0.8963*	*not rej.*	*not rej.*	*not rej.*	–	***1043.7043***	*0.9952*
	W	0.1369	not rej.	rej.	not rej.	–	869.421	0.9616	W	0.9071	not rej.	rej.	not rej.	–	1054.4661	0.9649
	B	0.1052	–	–	–	–	860.5262	0.992	B	0.8703	–	–	–	–	1050.4366	0.9964
Coimbra	LN	0.0005	not rej.	–	–	rej.	**932.6569**	0.9864	LN	*0.0948*	*not rej.*	–	–	*rej.*	***1026.203***	*0.9874*
	G	0	rej.	rej.	rej.	–	946.437	0.9796	G	0.0053	rej.	rej.	rej.	–	1041.9676	0.9625
	W	0	rej.	rej.	rej.	–	980.073	0.8959	W	0	rej.	rej.	rej.	–	1082.897	0.8988
	B	*0.0011*	–	–	–	–	*941.1955*	*0.9676*	B	0.216	–	–	–	–	1028.1304	0.9844

Tab. B.6: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Alcácer do Sal, Alcobaça, Barreiro, Beja, Cascais, Castro Verde, Chaves and Coimbra, in Winter. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Winter															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Alcácer do Sal	LN	0.1363	not rej.	–	–	rej.	1005.6851	0.9857	LN	0.2989	not rej.	–	–	not rej.	1050.3699	0.9923
	G	*0.2095*	*not rej.*	*not rej.*	*not rej.*	–	***1002.0219***	*0.9891*	G	*0.4208*	*not rej.*	*not rej.*	*not rej.*	–	***1049.1246***	*0.9926*
	W	0.0486	not rej.	rej.	rej.	–	1013.4445	0.9642	W	0.0784	not rej.	rej.	rej.	–	1063.3434	0.9581
	B	0.129	–	–	–	–	1009.2929	0.9857	B	0.3184	–	–	–	–	1058.0414	0.9876
Alcobaça	LN	*0.056*	*not rej.*	–	–	*not rej.*	***853.7245***	*0.9956*	LN	0.0372	rej.	–	–	rej.	1121.8564	0.977
	G	0.0496	not rej.	not rej.	not rej.	–	855.8276	0.9933	G	0.1358	not rej.	not rej.	not rej.	–	**1112.3181**	0.9892
	W	0.0006	rej.	rej.	rej.	–	878.485	0.9546	W	*0.1162*	*not rej.*	*not rej.*	*not rej.*	–	*1114.872*	*0.9837*
	B	0.0199	–	–	–	–	861.8872	0.9887	B	0.0869	–	–	–	–	1114.2915	0.9941
Barreiro	LN	0.5563	not rej.	–	–	rej.	958.053	0.9844	LN	*0.0288*	*not rej.*	–	–	*not rej.*	***1070.512***	*0.9943*
	G	*0.78*	*not rej.*	*not rej.*	*not rej.*	–	***952.314***	*0.9964*	G	0.0427	not rej.	not rej.	not rej.	–	1075.6968	0.9921
	W	0.1869	not rej.	rej.	rej.	–	966.7921	0.9731	W	0.0022	rej.	rej.	rej.	–	1100.5836	0.9323
	B	0.613	–	–	–	–	956.8583	0.9937	B	0.0068	–	–	–	–	1083.6683	0.9874
Beja	LN	*0.5186*	*not rej.*	–	–	*not rej.*	***1028.4607***	*0.9921*	LN	*0.0254*	*not rej.*	–	–	*rej.*	***1012.5884***	*0.9908*
	G	0.281	not rej.	rej.	rej.	–	1039.8813	0.9876	G	0.0061	not rej.	rej.	rej.	–	1018.875	0.9914
	W	0.0073	rej.	rej.	rej.	–	1071.0068	0.9112	W	0	rej.	rej.	rej.	–	1044.7954	0.921
	B	0.1389	–	–	–	–	1041.1859	0.9822	B	0.0023	–	–	–	–	1028.3023	0.9731
Cascais	LN	*0.1485*	*not rej.*	–	–	*not rej.*	***1048.6849***	*0.9951*	LN	0.2152	not rej.	–	–	not rej.	1093.2083	0.9883
	G	0.1261	not rej.	not rej.	not rej.	–	1053.5316	0.9903	G	0.3426	not rej.	not rej.	not rej.	–	**1089.5898**	0.9923
	W	0.0035	rej.	rej.	rej.	–	1080.1222	0.9331	W	0.0252	rej.	rej.	rej.	–	1107.6912	0.9732
	B	0.018	–	–	–	–	1060.8767	0.9852	B	*0.3781*	–	–	–	–	*1092.3411*	*0.9957*
Castro Verde	LN	0.5092	not rej.	–	–	rej.	982.1375	0.9725	LN	0.558	not rej.	–	–	rej.	1093.8225	0.9915
	G	*0.7973*	*not rej.*	*not rej.*	*not rej.*	–	***973.8514***	*0.9904*	G	*0.6396*	*not rej.*	*not rej.*	*not rej.*	–	***1093.4846***	*0.9916*
	W	0.3964	not rej.	rej.	rej.	–	984.4912	0.9629	W	0.1454	rej.	rej.	rej.	–	1107.8946	0.9508
	B	0.5297	–	–	–	–	979.1447	0.9911	B	0.4359	–	–	–	–	1103.9864	0.9857
Chaves	LN	0.1627	not rej.	–	–	rej.	917.341	0.9263	LN	0.5925	not rej.	–	–	rej.	1028.0439	0.9871
	G	*0.6207*	*not rej.*	*not rej.*	*not rej.*	–	***893.2932***	*0.9931*	G	*0.7578*	*not rej.*	*not rej.*	*not rej.*	–	***1021.009***	*0.9962*
	W	0.4058	not rej.	rej.	not rej.	–	897.0925	0.9533	W	0.2523	not rej.	rej.	rej.	–	1028.9318	0.9837
	B	0.4741	–	–	–	–	894.7471	0.9942	B	0.5794	–	–	–	–	1026.3508	0.9931
Coimbra	LN	0.034	not rej.	–	–	rej.	**913.6322**	0.9853	LN	*0.8427*	*not rej.*	–	–	*not rej.*	***984.5724***	*0.9966*
	G	*0.0218*	*rej.*	*rej.*	*rej.*	–	*917.9028*	*0.9899*	G	0.7183	not rej.	not rej.	not rej.	–	989.2514	0.9781
	W	0.0006	rej.	rej.	rej.	–	037.4016	0.919	W	0.0352	rej.	rej.	rej.	–	1016.7571	0.9413
	B	0.0065	–	–	–	–	929.9803	0.9702	B	0.6445	–	–	–	–	994.6834	0.985

Tab. B.7: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Alcácer do Sal, Alcobaça, Barreiro, Beja, Cascais, Castro Verde, Chaves and Coimbra, in Spring. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Spring															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Alcácer do Sal	LN	0	rej.	–	–	rej.	907.5194	0.967	LN	0.0559	not rej.	–	–	rej.	905.4432	0.9748
	G	0.0006	rej.	rej.	rej	–	897.0653	0.9842	G	*0.1512*	*not rej.*	*not rej.*	*not rej.*	–	*897.8961*	*0.9918*
	W	0.0045	not rej.	rej.	rej.	–	903.7601	0.9858	W	0.0671	not rej.	rej.	rej.	–	899.8104	0.9885
	B	*0.018*	–	–	–	–	***891.4343***	*0.988*	B	0.3024	–	–	–	–	**894.954**	0.9923
Alcobaça	LN	0	rej.	–	–	rej.	705.5636	0.8706	LN	0.2527	not rej.	–	–	not rej.	**1004.3754**	0.9961
	G	0	rej.	rej.	rej.	–	687.591	0.9617	G	*0.2906*	*not rej.*	*not rej.*	*not rej.*	–	*1006.7337*	*0.9871*
	W	0	rej.	rej.	rej.	–	713.5641	0.8712	W	0.0083	rej.	rej.	rej.	–	1038.4363	0.939
	B	*0.0019*	–	–	–	–	***662.9138***	*0.9822*	B	0.0789	–	–	–	–	1015.4609	0.9868
Barreiro	LN	*0.851*	*not rej.*	–	–	*not rej.*	***824.0675***	*0.9957*	LN	*0.1556*	*not rej.*	–	–	*not rej.*	***978.9125***	*0.9935*
	G	0.5344	rej.	rej.	rej.	–	828.6578	0.9909	G	0.0962	not rej.	not rej.	not rej.	–	983.303	0.9649
	W	0.0001	rej.	rej.	rej.	–	869.6605	0.9276	W	0	rej.	rej.	rej.	–	1030.9061	0.9326
	B	0.7341	–	–	–	–	831.8246	0.9857	B	0.0817	–	–	–	–	982.1248	0.9937
Beja	LN	0	rej.	–	–	rej.	922.1721	0.9809	LN	0.2061	rej.	–	–	rej.	982.706	0.9752
	G	0.0003	rej.	rej.	rej.	–	**917.1209**	0.9863	G	0.5844	not rej.	rej.	rej.	–	973.53	0.9881
	W	0	rej.	rej.	rej.	–	931.8229	0.9707	W	*0.9309*	*not rej.*	*not rej.*	*not rej.*	–	*969.791*	*0.9891*
	B	*0.0008*	–	–	–	–	*918.4885*	0.987	B	0.917	–	–	–	–	**969.4356**	0.9953
Cascais	LN	0.0004	rej.	–	–	rej.	1069.4464	0.9785	LN	0.0799	rej.	–	–	rej.	1179.5815	0.9679
	G	0.0029	not rej.	rej.	rej.	–	1060.6084	0.9878	G	0.482	not rej.	rej.	rej.	–	1167.2349	0.9684
	W	*0.0075*	*not rej.*	*not rej.*	*not rej.*	–	*1060.547*	*0.9842*	W	*0.9719*	*not rej.*	*not rej.*	*not rej.*	–	***1157.2747***	*0.9904*
	B	0.0058	–	–	–	–	**1059.955**	0.9934	B	0.9472	–	–	–	–	1159.5391	0.993
Castro Verde	LN	0.0173	rej.	–	–	not rej.	873.5506	0.9904	LN	0.0532	not rej.	–	–	rej.	1012.2096	0.9803
	G	0.0357	rej.	not rej.	not rej.	–	**869.9516**	0.9929	G	0.1438	not rej.	not rej.	not rej.	–	**1004.7424**	0.9902
	W	0.0007	rej.	rej.	rej.	–	885.8472	0.9722	W	0.0841	not rej.	rej.	rej.	–	1007.5435	0.9857
	B	*0.0477*	–	–	–	–	874.654	*0.9921*	B	*0.1457*	–	–	–	–	1004.8553	*0.9947*
Chaves	LN	0.0019	rej.	–	–	rej.	927.7293	0.9609	LN	0.0048	rej.	–	–	rej.	1038.1839	0.9429
	G	0.0508	not rej.	rej.	not rej.	–	912.8663	0.9938	G	0.0716	not rej.	rej.	rej.	–	1020.4372	0.9891
	W	0.2495	not rej.	rej.	rej.	–	910.4946	0.989	W	0.5434	not rej.	not rej.	not rej.	–	1006.7543	0.9937
	B	*0.2979*	–	–	–	–	***908.3318***	*0.9959*	B	*0.5749*	–	–	–	–	***1005.298***	*0.9885*
Coimbra	LN	0.0121	rej.	–	–	rej.	747.3983	0.9815	LN	0.044	not rej.	–	–	rej.	781.6078	0.9817
	G	0.004	rej.	rej.	rej.	–	753.4926	0.9408	G	0.0335	rej.	rej.	rej.	–	785.9571	0.9239
	W	0	rej.	rej.	rej.	–	806.9389	0.9224	W	0	rej.	rej.	rej.	–	843.7685	0.9268
	B	*0.0862*	–	–	–	–	***740.8824***	*0.9936*	B	*0.0938*	–	–	–	–	***776.1258***	*0.9892*

Tab. B.8: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Alcácer do Sal, Alcobaça, Barreiro, Beja, Cascais, Castro Verde, Chaves and Coimbra, in Summer. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Summer															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Alcácer do Sal	LN	0.0076	not rej.	–	–	rej.	662.008	0.9788	LN	0.0005	not rej.	–	–	rej.	748.5534	0.9637
	G	*0.0338*	*rej.*	*rej.*	*rej.*	–	*656.2876*	*0.9923*	G	0.0029	not rej.	rej.	rej.	–	741.9247	0.9739
	W	0.154	rej.	rej.	rej.	–	6611.303	0.9867	W	0.0385	not rej.	rej.	rej.	–	745.3621	0.9852
	B	0.2591	–	–	–	–	**653.156**	0.9925	B	*0.1062*	–	–	–	–	***731.4176***	*0.9795*
Alcobaça	LN	0.0145	rej.	–	–	rej.	553.5074	0.9096	LN	0.2738	not rej.	–	–	not rej.	935.5981	0.9936
	G	0.0648	rej.	rej.	rej.	–	540.2612	0.9413	G	*0.378*	*not rej.*	*not rej.*	*not rej.*	–	***934.5304***	*0.9914*
	W	0.015	rej.	rej.	rej.	–	550.7167	0.9512	W	0.032	not rej.	rej.	rej.	–	956.4474	0.9589
	B	*0.5883*	–	–	–	–	***516.7039***	*0.9674*	B	0.2263	–	–	–	–	942.0155	0.9931
Barreiro	LN	0.3079	rej.	–	–	rej.	817.7288	0.9071	LN	*0.1745*	*not rej.*	–	–	*not rej.*	***999.6312***	*0.9922*
	G	0.5993	not rej.	not rej.	not rej.	–	798.7288	0.9918	G	0.0871	rej.	rej.	rej.	–	1003.2183	0.9915
	W	0.0445	rej.	rej.	rej.	–	799.7356	0.9359	W	0.0001	rej.	rej.	rej.	–	1035.071	0.9246
	B	*0.6149*	–	–	–	–	***785.4964***	*0.9954*	B	0.0386	–	–	–	–	1012.6921	0.9621
(MAL) Beja	LN	0.315	not rej.	–	–	rej.	729.6217	0.9852	LN	0.0015	rej.	–	–	rej.	836.4962	0.935
	G	0.3879	not rej.	not rej.	not rej.	–	**725.7109**	0.9904	G	0.0168	rej.	rej.	rej.	–	822.7806	0.9594
	W	0.0018	rej.	rej.	rej.	–	741.9102	0.978	W	*0.9294*	*not rej.*	*not rej.*	*not rej.*	–	***791.0262***	*0.9964*
	B	*0.5365*	–	–	–	–	*725.8*	0.9904	B	0.8874	–	–	–	–	793.1539	0.9967
Cascais	LN	0.0014	rej.	–	–	rej.	1123.632	0.9367	LN	0.0119	rej.	–	–	rej.	1124.087	0.9598
	G	0.0259	rej.	rej.	rej.	–	1103.6016	0.9275	G	0.0775	rej.	rej.	rej.	–	1111.5921	0.9503
	W	*0.2086*	*not rej.*	*not rej.*	*not rej.*	–	***1078.3544***	*0.9901*	W	*0.4242*	*not rej.*	*not rej.*	*not rej.*	–	***1095.6024***	*0.9935*
	B	0.142	–	–	–	–	1081.269	0.9784	B	0.3373	–	–	–	–	1097.9247	0.9917
Castro Verde	LN	0.0246	rej.	–	–	rej.	699.3346	0.9796	LN	0.0265	rej.	–	–	rej.	824.2731	0.9572
	G	0.0962	not rej.	rej.	rej.	–	694.5093	0.9816	G	0.1104	not rej.	rej.	rej.	–	814.5785	0.9857
	W	0.1272	rej.	rej.	rej.	–	709.1051	0.9821	W	0.7325	not rej.	not rej.	not rej.	–	800.6751	0.9948
	B	*0.5292*	–	–	–	–	***690.7316***	*0.99*	B	*0.7323*	–	–	–	–	***799.8591***	*0.9931*
Chaves	LN	0.0768	rej.	–	–	rej.	814.0603	0.9692	LN	0.007	not rej.	–	–	rej.	900.8183	0.979
	G	0.2712	not rej.	not rej.	not rej.	–	804.0624	0.9659	G	0.0765	not rej.	not rej.	not rej.	–	893.0679	0.9891
	W	0.0759	not rej.	rej.	rej.	–	815.0381	0.987	W	0.5845	not rej.	not rej.	not rej.	–	895.3617	0.9883
	B	*0.3623*	–	–	–	–	***800.2199***	*0.9814*	B	*0.5438*	–	–	–	–	***891.3596***	*0.9914*
Coimbra	LN	*0.2057*	*not rej.*	–	–	*not rej.*	*550.2613*	*0.9942*	LN	0.8859	not rej.	–	–	rej.	703.4782	0.9693
	G	0.2308	not rej.	not rej.	not rej.	–	**549.867**	0.9929	G	*0.9383*	*not rej.*	*not rej.*	*not rej.*	–	*697.4881*	*0.997*
	W	0.0019	rej.	rej.	rej.	–	574.4337	0.9543	W	0.1206	not rej.	rej.	rej.	–	708.1481	0.9707
	B	0.3	–	–	–	–	556.2206	0.9904	B	0.8732	–	–	–	–	**695.5394**	0.9962

Tab. B.9: Fitted distributions to the **observed wind** of 8 stations: Alcácer do Sal, Alcobaça, Barreiro, Beja, Cascais, Castro Verde, Chaves and Coimbra. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn			Winter			Spring			Summer		
	$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\alpha}_{AS}$	5.0229	(4.2809,6.0926)	$\hat{\alpha}_{AS}$	4.458	(3.7728,5.4031)	$\hat{k}_{AS}$	2.5438	(1.423,11.2798)	$\hat{\alpha}_{AS}$	40.4841	(34.4933,48.9428)
$\hat{\beta}_{AS}$	0.9818	(0.8311,1.2013)	$\hat{\beta}_{AS}$	0.9035	(0.7593,1.1032)	$\hat{c}_{AS}$	5.3894	(4.5302,6.4768)	$\hat{\beta}_{AS}$	6.6625	(5.6672,8.0668)
–	–	–	–	–	–	$\hat{\lambda}_{AS}$	0.1327	(0.0903,0.1553)	–	–	–
$\hat{k}_{A}$	0.8807	(0.5858,1.4652)	$\hat{\mu}_{A}$	1.4311	(1.3807,1.4785)	$\hat{k}_{A}$	0.91	(0.6003,1.5226)	$\hat{k}_{A}$	1.8279	(1.0686,4.7134)
$\hat{c}_{A}$	5.5196	(4.5753,6.8496)	$\hat{\sigma}_{A}$	0.379	(0.3429,0.412)	$\hat{c}_{A}$	9.4844	(7.9107,11.7017)	$\hat{c}_{A}$	9.9587	(8.3264,12.1705)
$\hat{\lambda}_{A}$	0.2594	(0.221,0.2909)	–	–	–	$\hat{\lambda}_{A}$	0.2131	(0.1948,0.2271)	$\hat{\lambda}_{A}$	0.1948	(0.169,0.2111)
$\hat{\mu}_{Ba}$	1.6532	(1.609,1.6958)	$\hat{\alpha}_{Ba}$	7.8442	(6.654,9.4983)	$\hat{\mu}_{Ba}$	1.7929	(1.7653,1.8201)	$\hat{k}_{Ba}$	2.005	(1.1846,5.7481)
$\hat{\sigma}_{Ba}$	0.338	(0.3065,0.3672)	$\hat{\beta}_{Ba}$	1.3972	(1.1791,1.7055)	$\hat{\sigma}_{Ba}$	0.2193	(0.1989,0.2388)	$\hat{c}_{Ba}$	6.976	(5.8676,8.3548)
–	–	–	–	–	–	–	–	–	$\hat{\lambda}_{Ba}$	0.1472	(0.1175,0.1636)
$\hat{\mu}_{Be}$	1.7169	(1.6659,1.7669)	$\hat{\mu}_{Be}$	1.6802	(1.6268,1.7336)	$\hat{k}_{Be}$	2.0253	(1.1763,5.8077)	$\hat{k}_{Be}$	1.7295	(1.0382,4.2413)
$\hat{\sigma}_{Be}$	0.3963	(0.3595,0.432)	$\hat{\sigma}_{Be}$	0.4102	(0.3715,0.4453)	$\hat{c}_{Be}$	5.8269	(4.9091,7.0548)	$\hat{c}_{Be}$	9.5163	(7.9838,11.5872)
–	–	–	–	–	–	$\hat{\lambda}_{Be}$	0.1318	(0.1012,0.1507)	$\hat{\lambda}_{Be}$	0.1352	(0.1174,0.1464)
$\hat{\alpha}_{C}$	7.3518	(6.1641,9.0168)	$\hat{\mu}_{C}$	2.0246	(1.9787,2.0719)	$\hat{\omega}_{C}$	4.1565	(3.7939,4.6324)	$\hat{\omega}_{C}$	3.9007	(3.5551,4.3342)
$\hat{\beta}_{C}$	0.9446	(0.7913,1.1635)	$\hat{\sigma}_{C}$	0.3396	(0.3071,0.3698)	$\hat{\delta}_{C}$	9.1605	(8.8669,9.4607)	$\hat{\delta}_{C}$	9.2308	(8.913,9.5496)
$\hat{\alpha}_{CV}$	7.4713	(6.3241,9.0433)	$\hat{\alpha}_{CV}$	7.0543	(6.0126,8.5215)	$\hat{k}_{CV}$	1.9945	(1.1474,5.8637)	$\hat{k}_{CV}$	2.0952	(1.2037,6.6739)
$\hat{\beta}_{CV}$	1.3236	(1.1121,1.6136)	$\hat{\beta}_{CV}$	1.2811	(1.0849,1.553)	$\hat{c}_{CV}$	6.2969	(5.2858,7.6131)	$\hat{c}_{CV}$	9.3144	(7.8448,11.253)
–	–	–	–	–	–	$\hat{\lambda}_{CV}$	0.1363	(0.1068,0.1544)	$\hat{\lambda}_{CV}$	0.1364	(0.114,0.1483)
$\hat{\alpha}_{Ch}$	4.5478	(3.8467,5.5375)	$\hat{\alpha}_{Ch}$	3.5886	(3.0378,4.3628)	$\hat{k}_{Ch}$	3.9831	(1.8697,68.0279)	$\hat{k}_{Ch}$	2.4323	(1.3308,11.3605)
$\hat{\beta}_{Ch}$	1.1459	(0.9584,1.4061)	$\hat{\beta}_{Ch}$	0.8529	(0.7134,1.0554)	$\hat{c}_{Ch}$	4.4238	(3.786,5.2816)	$\hat{c}_{Ch}$	5.0856	(4.2557,6.1444)
–	–	–	–	–	–	$\hat{\lambda}_{Ch}$	0.1262	(0.0573,0.1586)	$\hat{\lambda}_{Ch}$	0.1609	(0.1057,0.1896)
$\hat{k}_{Co}$	0.6151	(0.3884,0.9197)	$\hat{\alpha}_{Co}$	7.1571	(6.0376,8.7257)	$\hat{k}_{Co}$	0.7918	(0.5181,1.3076)	$\hat{\mu}_{Co}$	1.701	(1.6813,1.7217)
$\hat{c}_{Co}$	5.6521	(4.4834,7.0615)	$\hat{\beta}_{Co}$	1.3699	(1.1476,1.6668)	$\hat{c}_{Co}$	8.9021	(7.3176,11.2098)	$\hat{\sigma}_{Co}$	0.152	(0.1378,0.1655)
$\hat{\lambda}_{Co}$	0.196	(0.2184,0.2824)	–	–	–	$\hat{\lambda}_{Co}$	0.1935	(0.176,0.2078)	–	–	–

Tab. B.10: Fitted distributions to the **simulated wind** of 8 stations: Alcácer do Sal, Alcobaça, Barreiro, Beja, Cascais, Castro Verde, Chaves and Coimbra. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn			Winter			Spring			Summer		
	$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\alpha}_{AS}$	6.9088	(5.8609, 8.415)	$\hat{\alpha}_{AS}$	6.9058	(5.8588, 8.3993)	$\hat{\alpha}_{AS}$	29.0966	(24.656, 35.2288)	$\hat{k}_{AS}$	2.668	(1.4478, 13.7433)
$\hat{\beta}_{AS}$	1.0577	(0.8857, 1.2975)	$\hat{\beta}_{AS}$	1.0441	(0.8805, 1.2761)	$\hat{\beta}_{AS}$	3.3694	(2.8533, 4.0811)	$\hat{c}_{AS}$	10.6627	(9.0431, 12.8043)
–	–	–	–	–	–	–	–	–	$\hat{\lambda}_{AS}$	0.1048	(0.0852, 0.1136)
$\hat{\alpha}_{A}$	9.0092	(7.6505, 10.8596)	$\hat{\omega}_{A}$	3.4169	(3.11, 3.8155)	$\hat{\alpha}_{A}$	19.2705	(16.2204, 23.3784)	$\hat{\alpha}_{A}$	21.8743	(18.4207, 26.5793)
$\hat{\beta}_{A}$	1.075	(0.9083, 1.3006)	$\hat{\delta}_{A}$	9.9398	(9.5358, 10.3415)	$\hat{\beta}_{A}$	2.2266	(1.8715, 2.7135)	$\hat{\beta}_{A}$	2.6059	(2.194, 3.1823)
$\hat{\mu}_{Ba}$	1.874	(1.8249, 1.9196)	$\hat{\mu}_{Ba}$	1.8819	(1.8321, 1.9302)	$\hat{\mu}_{Ba}$	2.1463	(2.12, 2.1729)	$\hat{\mu}_{Ba}$	2.1732	(2.1451, 2.2018)
$\hat{\sigma}_{Ba}$	0.3704	(0.3353, 0.4029)	$\hat{\sigma}_{Ba}$	0.3782	(0.3427, 0.4117)	$\hat{\sigma}_{Ba}$	0.212	(0.1923, 0.2304)	$\hat{\sigma}_{Ba}$	0.223	(0.203, 0.2421)
$\hat{\mu}_{Be}$	1.7404	(1.6867, 1.7914)	$\hat{\mu}_{Be}$	1.7799	(1.7332, 1.8259)	$\hat{\omega}_{Be}$	4.9215	(4.4946, 5.4731)	$\hat{\omega}_{Be}$	7.2944	(6.6375, 8.1131)
$\hat{\sigma}_{Be}$	0.4135	(0.3756, 0.449)	$\hat{\sigma}_{Be}$	0.3587	(0.3262, 0.3907)	$\hat{\delta}_{Be}$	8.6256	(8.3891, 8.8536)	$\hat{\delta}_{Be}$	8.3718	(8.2136, 8.5273)
$\hat{\alpha}_{C}$	7.9966	(6.7278, 9.7209)	$\hat{k}_{C}$	1.6194	(0.9495, 3.8033)	$\hat{\omega}_{C}$	4.1032	(3.7335, 4.5674)	$\hat{\omega}_{C}$	4.8355	(4.4028, 5.3622)
$\hat{\beta}_{C}$	0.9181	(0.7672, 1.1233)	$\hat{c}_{C}$	4.557	(3.7466, 5.5724)	$\hat{\delta}_{C}$	10.9439	(10.5832, 11.3056)	$\hat{\delta}_{C}$	11.704	(11.3592, 12.0276)
–	–	–	$\hat{\lambda}_{C}$	0.1028	(0.07703, 0.1221)	–	–	–	–	–	–
$\hat{\mu}_{CV}$	1.8614	(1.8133, 1.9081)	$\hat{\alpha}_{CV}$	6.8667	(5.8405, 8.3174)	$\hat{k}_{CV}$	3.5599	(1.8077, 64.8482)	$\hat{k}_{CV}$	5.5101	(2.2181, 140.8124)
$\hat{\sigma}_{CV}$	0.3665	(0.3334, 0.3978)	$\hat{\beta}_{CV}$	0.9741	(0.8194, 1.1855)	$\hat{c}_{CV}$	5.9392	(5.0351, 7.0601)	$\hat{c}_{CV}$	8.3708	(7.2548, 9.967)
–	–	–	–	–	–	$\hat{\lambda}_{CV}$	0.0882	(0.0477, 0.1034)	$\hat{\lambda}_{CV}$	0.0911	(0.058, 0.1049)
$\hat{\alpha}_{Ch}$	4.8197	(4.0609, 5.8692)	$\hat{\alpha}_{Ch}$	4.8387	(4.0895, 5.8943)	$\hat{k}_{Ch}$	5.4749	(2.2457, 118.7746)	$\hat{k}_{Ch}$	3.6515	(1.7814, 70.3708)
$\hat{\beta}_{Ch}$	0.7686	(0.6459, 0.941)	$\hat{\beta}_{Ch}$	0.7507	(0.6283, 0.9157)	$\hat{c}_{Ch}$	5.2531	(4.5339, 6.2247)	$\hat{c}_{Ch}$	5.777	(4.9002, 6.9059)
–	–	–	–	–	–	$\hat{\lambda}_{Ch}$	0.0846	(0.0424, 0.1056)	$\hat{\lambda}_{Ch}$	0.1023	(0.0539, 0.1219)
$\hat{\mu}_{Co}$	1.7928	(1.7424, 1.8417)	$\hat{\mu}_{Co}$	1.8001	(1.7536, 1.8496)	$\hat{k}_{Co}$	0.9012	(0.5893, 1.5681)	$\hat{\alpha}_{Co}$	37.0076	(30.9652, 45.0857)
$\hat{\sigma}_{Co}$	0.3828	(0.3471, 0.4186)	$\hat{\sigma}_{Co}$	0.3567	(0.3228, 0.3881)	$\hat{c}_{Co}$	10.4629	(8.6061, 13.1563)	$\hat{\beta}_{Co}$	5.1884	(4.327, 6.3267)
–	–	–	–	–	–	$\hat{\lambda}_{Co}$	0.1403	(0.1282, 0.149)	–	–	–

Tab. B.11: Copulas selected according to the AIC to fit 8 stations: Alcácer do Sal, Alcobaça, Barreiro, Beja, Cascais, Castro Verde, Chaves and Coimbra. $\hat{\theta}_1$ and $\hat{\theta}_2$ are the copula parameters estimates obtained by MPLE.

Station	Season	Copula	$\hat{\theta}_{MPLE}$				Tail Dependence [1]		p-values		
			$\hat{\theta}_1$	CI (95%)	$\hat{\theta}_2$	CI (95%)	$\hat{\lambda}_L$	$\hat{\lambda}_U$	$S_n^{(B)}$	S_n	$S_n^{(K)}$
Alcácer do Sal	Autumn	C_α^{gu}	2.3725	(2.0721, 2.6728)	–	–	0	0.6607	0.872	1	0.49
	Winter	C_α^{gu}	2.2851	(2.0605, 2.5097)	–	–	0	0.6456	0.977	0.999	0.66
	Spring	C_α^{gu}	1.5727	(1.403, 1.7425)	–	–	0	0.4462	0.85	1	0.95
	Summer	C_α^{sg}	1.2448	(1.1208, 1.3689)	–	–	0.2549	0	0.785	1	0.59
Alcobaça	Autumn	C_α^{j}	1.7496	(1.4393, 2.0599)	–	–	0	0.5139	(**)	0.0125	0.8
	Winter	C_α^{gu}	1.5925	(1.393, 1.7921)	–	–	0	0.4547	0.404	0.906	0.72
	Spring	$C_{\rho\eta}^{t}$	0.4823	(0.3808, 0.5829)	4.053	(*)	0.2409	0.2409	0.277	0.0445	0.29
	Summer	$C_{\rho\eta}^{t}$	0.4575	(0.3444, 0.5707)	3.7588	(*)	0.2434	0.2434	0.391	0.4431	0.47
Barreiro	Autumn	C_α^{gu}	2.3003	(2.0149, 2.5857)	–	–	0	0.6483	0.786	0.993	0.67
	Winter	C_α^{f}	7.7577	(6.5318, 8.9836)	–	–	0	0	0.475	1	0.38
	Spring	C_α^{sc}	1.3832	(3.6432, 5.6347)	–	–	0	0.6059	0.249	0.098	0.73
	Summer	C_α^{gu}	1.6934	(1.4835, 1.9033)	–	–	0	0.4942	0.475	1	0.38
Beja	Autumn	C_α^{f}	8.0217	(6.6447, 9.3987)	–	–	0	0	0.807	0.988	0.45
	Winter	C_α^{gu}	2.1967	(1.9406, 2.4529)	–	–	0	0.629	0.997	0.997	0.63
	Spring	C_ρ^{gaus}	0.7049	(0.6416, 0.7682)	–	–	0	0	0.982	1	0.67
	Summer	C_α^{c}	0.8366	(0.5244, 1.1488)	–	–	0.4367	0	0.918	1	0.3
Cascais	Autumn	C_α^{f}	8.6497	(7.2207, 10.0788)	–	–	0	0	0.766	1	0.75
	Winter	C_α^{gu}	2.3426	(2.0691, 2.616)	–	–	0	0.6557	0.92	0.989	0.68
	Spring	C_α^{f}	7.4946	(6.0794, 8.9098)	–	–	0	0	0.882	0.993	0.35
	Summer	C_ρ^{gaus}	0.8075	(0.7675, 0.8476)	–	–	0	0	0.921	1	0.9
Castro Verde	Autumn	C_α^{gu}	2.2682	(1.9651, 2.5714)	–	–	0	0.6426	0.929	1	0.59
	Winter	C_α^{gu}	2.56	(2.2368, 2.8831)	–	–	0	0.689	0.884	0.975	0.78
	Spring	C_α^{gu}	1.9768	(1.7443, 2.2092)	–	–	0	0.58	0.927	0.998	0.81
	Summer	C_ρ^{gaus}	0.4994	(0.4177, 0.581)	–	–	0	0	0.995	1	0.84
Chaves	Autumn	$C_{\rho\eta}^{t}$	0.7493	(0.6829, 0.8156)	4.8174	(*)	0.3975	0.3975	0.999	0.1873	0.04
	Winter	C_α^{gu}	2.3108	(2.0106, 2.6111)	–	–	0	0.6502	0.978	0.999	0.57
	Spring	$C_{\rho\eta}^{t}$	0.6292	(0.5551, 0.7034)	9.0188	(*)	0.1619	0.1619	0.855	0.6009	0.79
	Summer	C_α^{f}	3.7898	(2.8763, 4.7033)	–	–	0	0	0.981	1	0.58
Coimbra	Autumn	C_α^{gu}	2.5595	(2.3085, 2.8105)	–	–	0	0.689	0.667	1	0.53
	Winter	C_α^{f}	7.4736	(6.0441, 8.9032)	–	–	0	0	0.607	1	0.6
	Spring	$C_{\rho\eta}^{t}$	0.6423	(0.5644, 0.7202)	5.381	(*)	0.2805	0.2805	0.83	0.1973	0
	Summer	C_α^{f}	5.0671	(3.9902, 6.144)	–	–	0	0	0.657	1	0.87

(*) At present the asymptotic variance cannot be fully estimated if η is not fixed, thus it is not possible to provide a confidence interval; see the package *copula* manual.
(**) $S_n^{(B)}$ is not implemented for the Joe Copula; see the package *gofCopula* manual. [1] the 0's are theoretical.

Tab. B.12: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Elvas, Évora, Faro, Lisboa S2, Lousã, Manteigas, Matosinhos and Mértola, in Autumn. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Autumn															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Elvas	LN	0.1727	not rej.	–	–	not rej.	846.5635	0.994	LN	*0.1392*	*not rej.*	–	–	*not rej.*	***923.6404***	*0.9942*
	G	*0.4422*	*not rej.*	*not rej.*	*not rej.*	–	***846.4701***	*0.9825*	G	0.0906	not rej.	not rej.	not rej.	–	928.8001	0.982
	W	0.1023	rej.	rej.	rej.	–	867.0939	0.9635	W	0.0012	rej.	rej.	rej.	–	954.9854	0.9399
	B	0.29	–	–	–	–	850.8513	0.992	B	0.0262	–	–	–	–	932.7272	0.9953
Évora	LN	*0.2686*	*not rej.*	–	–	*not rej.*	***1110.4788***	*0.9958*	LN	*0.596*	*not rej.*	–	–	*not rej.*	***1020.1568***	*0.9958*
	G	0.0414	not rej.	rej.	rej.	–	1114.6202	0.986	G	0.162	not rej.	rej.	rej.	–	1026.3002	0.995
	W	0	rej.	rej.	rej.	–	1137.6202	0.9503	W	0.0006	rej.	rej.	rej.	–	1051.9733	0.9361
	B	0.2426	–	–	–	–	1119.4238	0.9744	B	0.1572	–	–	–	–	1032.8686	0.9778
Faro	LN	*0.8873*	*not rej.*	–	–	*not rej.*	***1122.6177***	*0.9953*	LN	*0.1869*	*not rej.*	–	–	*not rej.*	***1162.1443***	*0.9963*
	G	0.7256	not rej.	rej.	not rej.	–	1132.5581	0.9699	G	0.1341	not rej.	not rej.	not rej.	–	1163.36	0.9922
	W	0.0058	rej.	rej.	rej.	–	1167.4156	0.9277	W	0.001	rej.	rej.	rej.	–	1186.2388	0.9584
	B	0.6509	–	–	–	–	1130.9203	0.9897	B	0.1805	–	–	–	–	1168.7626	0.991
Lisboa S2	LN	*0.0633*	*not rej.*	–	–	*not rej.*	***952.4061***	*0.9939*	LN	0.2263	not rej.	–	–	rej.	1072.9632	0.9752
	G	0.011	not rej.	rej.	rej.	–	958.3716	0.9725	G	0.4939	not rej.	not rej.	not rej.	–	**1062.2906**	0.992
	W	0	rej.	rej.	rej.	–	992.5647	0.9424	W	0.1201	rej.	rej.	rej.	–	1070.9954	0.9883
	B	0.0493	–	–	–	–	955.0443	0.9893	B	*0.5133*	–	–	–	–	*1062.547*	*0.9918*
Lousã	LN	0.0032	rej.	–	–	rej.	824.584	0.9696	LN	*0.074*	*not rej.*	–	–	*rej.*	***1088.1353***	*0.9862*
	G	0	rej	rej	rej	–	859.0831	0.8603	G	0.0043	rej.	rej.	rej.	–	1097.8405	0.9887
	W	0	rej.	rej.	rej.	–	908.2147	0.8652	W	0	rej.	rej.	rej.	–	1127.6148	0.9025
	B	*0.1496*	–	–	–	–	***811.701***	*0.9945*	B	0.0451	–	–	–	–	1100.9988	0.9541
Manteigas	LN	0.4924	not rej.	–	–	not rej.	1206.6293	0.9946	LN	0.8465	not rej.	–	–	not rej.	1067.4047	0.9936
	G	*0.5965*	*not rej.*	*not rej.*	*not rej.*	–	***1205.7914***	*0.9881*	G	*0.886*	*not rej.*	*not rej.*	*not rej.*	–	***1065.3648***	*0.9951*
	W	0.1717	not rej.	rej.	rej.	–	1221.0117	0.9659	W	0.2791	not rej.	rej.	rej.	–	1077.9337	0.9642
	B	0.4598	–	–	–	–	1212.2146	0.9925	B	0.6779	–	–	–	–	1073.5118	0.9913
Matosinhos	LN	*0.2719*	*not rej.*	–	–	*not rej.*	***1050.0692***	*0.9953*	LN	0.6908	not rej.	–	–	not rej.	1236.1073	0.9911
	G	0.146	not rej.	not rej.	not rej.	–	1054.5107	0.9788	G	*0.6977*	*not rej.*	*not rej.*	*not rej.*	–	***1234.116***	*0.9894*
	W	0.001	rej.	rej.	rej.	–	1082.2162	0.9414	W	0.1048	rej.	rej.	rej.	–	1248.9478	0.9635
	B	0.0570	–	–	–	–	1061.0777	0.9903	B	0.6961	–	–	–	–	1241.6371	0.9817
Mértola	LN	*0.6015*	*not rej.*	–	–	*not rej.*	***942.165***	*0.9953*	LN	*0.5258*	*not rej.*	–	–	*not rej.*	***1025.631***	*0.9942*
	G	0.1124	rej.	rej.	rej.	–	952.1767	0.9673	G	0.337	not rej.	rej.	rej.	–	1034.2286	0.9662
	W	0	rej.	rej.	rej.	–	987.6168	0.9302	W	0.0102	rej.	rej.	rej.	–	1063.6876	0.9273
	B	0.5629	–	–	–	–	947.8556	0.989	B	0.2377	–	–	–	–	1037.8187	0.9801

Tab. B.13: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Elvas, Évora, Faro, Lisboa S2, Lousã, Manteigas, Matosinhos and Mértola, in Winter. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Winter															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Elvas	LN	*0.7996*	*not rej.*	–	–	*rej.*	*876.1672*	*0.9632*	LN	*0.3897*	*not rej.*	–	–	*not rej.*	***949.4749***	*0.9917*
	G	0.2433	not rej.	not rej.	not rej.	–	**870.7384**	0.9869	G	0.1927	not rej.	rej.	rej.	–	953.4519	0.9918
	W	0.0032	rej.	rej.	rej.	–	886.1135	0.9445	W	0.007	rej.	rej.	rej.	–	970.4564	0.9356
	B	0.3738	–	–	–	–	873.7112	0.9808	B	0.0492	–	–	–	–	963.0652	0.974
Évora	LN	0.3366	not rej.	–	–	not rej.	**1091.6876**	0.995	LN	*0.3887*	*not rej.*	–	–	*not rej.*	***1027.6587***	*0.9957*
	G	*0.1547*	*rej.*	*rej.*	*rej.*	–	*1098.0617*	*0.9872*	G	0.3948	not rej.	not rej.	not rej.	–	1030.0988	0.9706
	W	0.0023	rej.	rej.	rej.	–	1123.0827	0.9352	W	0.0281	rej.	rej.	rej.	–	1055.0079	0.9538
	B	0.1375	–	–	–	–	1104.8012	0.9881	B	0.326	–	–	–	–	1034.7381	0.9935
Faro	LN	*0.01*	*not rej.*	–	–	*not rej.*	***1074.9819***	*0.9941*	LN	0.5064	not rej.	–	–	rej.	1099.12	0.9839
	G	0.0023	rej.	rej.	rej.	–	1079.7622	0.988	G	0.3438	not rej.	rej.	rej.	–	1096.5988	0.9894
	W	0	rej.	rej.	rej.	–	1108.5258	0.944	W	0.0007	rej.	rej.	rej.	–	1120.7235	0.9656
	B	0.0061	–	–	–	–	1081.387	0.9801	B	*0.6318*	–	–	–	–	***1095.8932***	*0.9907*
Lisboa S2	LN	*0.0752*	*not rej.*	–	–	*not rej.*	***965.1377***	*0.9942*	LN	0.1958	rej.	–	–	rej.	1089.4389	0.9836
	G	0.0331	not rej.	not rej.	not rej.	–	970.9786	0.9655	G	0.5642	not rej.	not rej.	not rej.	–	**1083.6092**	0.957
	W	0	rej.	rej.	rej.	–	1005.5485	0.9406	W	0.2332	not rej.	rej.	rej.	–	1099.4467	0.9773
	B	0.0473	–	–	–	–	969.3644	0.996	B	*0.6258*	–	–	–	–	*1085.4138*	*0.98*
Lousã	LN	*0.1816*	*not rej.*	–	–	*rej.*	***660.9826***	*0.9819*	LN	*0.2163*	*not rej.*	–	–	*rej.*	***912.9374***	*0.9851*
	G	0.0158	rej	rej	rej	–	676.5265	0.9801	G	0.1529	rej.	rej.	rej.	–	918.2807	0.9577
	W	0.0001	rej.	rej.	rej.	–	702.5402	0.8885	W	0.0064	rej.	rej.	rej.	–	943.065	0.918
	B	0.1816	–	–	–	–	665.3619	0.9601	B	0.0354	–	–	–	–	925.3281	0.9874
Manteigas	LN	0.0486	not rej.	–	–	not rej.	1238.5889	0.9899	LN	0.4562	not rej.	–	–	rej.	1097.0675	0.987
	G	*0.1662*	*not rej.*	*not rej.*	*not rej.*	–	***1235.7315***	*0.9866*	G	*0.7839*	*not rej.*	*not rej.*	*not rej.*	–	***1093.0615***	*0.9897*
	W	0.0544	rej.	rej.	rej.	–	1246.1655	0.9671	W	0.6134	not rej.	rej.	not rej.	–	1099.9316	0.9641
	B	0.0685	–	–	–	–	1243.2206	0.9894	B	0.6266	–	–	–	–	1100.4354	0.9959
Matosinhos	LN	*0.1722*	*not rej.*	–	–	*not rej.*	***942.0021***	*0.9914*	LN	*0.016*	*not rej.*	–	–	*not rej.*	*1145.3339*	*0.9886*
	G	0.0314	rej.	rej.	rej.	–	946.1897	0.9848	G	0.0172	not rej.	not rej.	not rej.	–	**1142.7641**	0.9839
	W	0	rej.	rej.	rej.	–	978.2768	0.9417	W	0.0001	rej.	rej.	rej.	–	1161.1943	0.9699
	B	0.2284	–	–	–	–	944.8086	0.9856	B	0.05	–	–	–	–	1145.9216	0.9817
Mértola	LN	*0.6265*	*not rej.*	–	–	*not rej.*	***915.4267***	*0.995*	LN	0.0729	not rej.	–	–	not rej.	**1011.4698**	0.995
	G	0.5938	rej.	not rej.	not rej.	–	917.2715	0.9891	G	*0.0949*	*not rej.*	*not rej.*	*not rej.*	–	*1913.7654*	0.9875
	W	0.0482	rej.	rej.	rej.	–	939.3792	0.9589	W	0.0056	not rej.	rej.	rej.	–	1035.3417	0.9503
	B	0.4913	–	–	–	–	921.5641	0.9917	B	0.0285	–	–	–	–	1020.6989	0.9962

Tab. B.14: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Elvas, Évora, Faro, Lisboa S2, Lousã, Manteigas, Matosinhos and Mértola, in Spring. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Spring															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Elvas	LN	0.0002	rej.	–	–	rej.	844.2566	0.9565	LN	0.2737	not rej.	–	–	not rej.	917.7532	0.9909
	G	0.0043	rej.	rej.	rej.	–	831.6393	0.9712	G	0.5879	not rej.	not rej.	not rej.	–	**916.3628**	0.9565
	W	*0.1628*	*not rej.*	*not rej.*	*rej.*	–	***816.7194***	*0.9909*	W	0.2295	not rej.	rej.	rej.	–	944.7552	0.9639
	B	0.1098	–	–	–	–	818.5554	0.9852	B	*0.6407*	–	–	–	–	*919.0273*	*0.9884*
Évora	LN	0.5211	not rej.	–	–	not rej.	979.754	0.9944	LN	0.0105	rej.	–	–	rej.	895.2368	0.9818
	G	*0.7821*	*not rej.*	*not rej.*	*not rej.*	–	***977.5608***	*0.993*	G	0.0487	not rej.	rej.	rej.	–	889.8005	0.9711
	W	0.1601	not rej.	rej.	rej.	–	996.3824	0.9667	W	0.0306	rej.	rej.	rej.	–	907.8512	0.9804
	B	0.7975	–	–	–	–	984.45	0.9898	B	*0.1151*	–	–	–	–	***887.9355***	*0.9879*
Faro	LN	0.0092	rej.	–	–	rej.	1034.4915	0.98	LN	0	rej.	–	–	rej.	1052.2011	0.9406
	G	0.0743	not rej.	rej.	rej.	–	**1025.9471**	0.9905	G	0	rej.	rej.	rej.	–	1035.1825	0.9708
	W	*0.2429*	*not rej.*	*not rej.*	*not rej.*	–	*1027.8608*	*0.9852*	W	*0.0238*	*not rej.*	*not rej.*	*not rej.*	–	***1010.7444***	*0.9965*
	B	0.2057	–	–	–	–	1026.1803	0.9948	B	0.0047	–	–	–	–	1016.9558	0.9919
Lisboa S2	LN	0.3871	not rej.	–	–	not rej.	822.2777	0.988	LN	*0.8344*	*not rej.*	–	–	*not rej.*	***1079.6414***	*0.9969*
	G	*0.444*	*not rej.*	*not rej.*	*not rej.*	–	*822.335*	*0.9413*	G	0.7659	not rej.	not rej.	not rej.	–	1080.1804	0.9966
	W	0.0006	rej.	rej.	rej.	–	863.4006	0.9569	W	0.0316	rej.	rej.	rej.	–	1103.7294	0.9531
	B	0.4215	–	–	–	–	**821.48**	0.9899	B	0.418	–	–	–	–	1090.3252	0.9873
Lousã	LN	0.0118	rej.	–	–	rej.	630.2853	0.9612	LN	0.362	not rej.	–	–	rej.	**794.2166**	0.9848
	G	0.0006	rej	rej	rej	–	642.8882	0.858	G	0.0828	not rej.	rej.	rej.	–	800.5197	0.9727
	W	0	rej.	rej.	rej.	–	702.2748	0.8819	W	0	rej.	rej.	rej.	–	848.9219	0.8968
	B	*0.0747*	–	–	–	–	***619.5451***	*0.9838*	B	*0.931*	–	–	–	–	*794.2375*	*0.9777*
Manteigas	LN	0.1411	rej.	–	–	rej.	1167.9829	0.9866	LN	0.5374	not rej.	–	–	not rej.	**938.5205**	0.9937
	G	*0.2383*	*not rej.*	*not rej.*	*not rej.*	–	***1165.6293***	*0.9792*	G	0.4024	not rej.	rej.	rej.	–	940.477	0.984
	W	0.034	rej.	rej.	rej.	–	1175.453	0.9543	W	0.0011	rej.	rej.	rej.	–	971.6921	0.9459
	B	0.1093	–	–	–	–	1174.8693	0.9806	B	*0.7989*	–	–	–	–	*942.0714*	*0.984*
Matosinhos	LN	0.5314	not rej.	–	–	not rej.	1016.6048	0.9904	LN	0.7625	not rej.	–	–	rej.	1167.3437	0.9907
	G	*0.8654*	*not rej.*	*not rej.*	*not rej.*	–	***1012.0484***	*0.9955*	G	*0.9615*	*not rej.*	*not rej.*	*not rej.*	–	***1163.0663***	*0.9938*
	W	0.6594	not rej.	rej.	not rej.	–	1024.971	0.9734	W	0.6685	not rej.	rej.	not rej.	–	1173.3655	0.9699
	B	0.8181	–	–	–	–	1017.6869	0.9976	B	0.9284	–	–	–	–	1170.3112	0.9944
Mértola	LN	0.0405	rej.	–	–	rej.	848.9033	0.9813	LN	0.2528	not rej.	–	–	rej.	911.5081	0.9836
	G	0.2826	not rej.	not rej.	not rej.	–	**841.8233**	0.9896	G	0.618	not rej.	rej.	rej.	–	**905.6604**	0.9848
	W	0.5417	not rej.	not rej.	not rej.	–	844.0914	0.9772	W	0.9116	not rej.	not rej.	not rej.	–	909.4131	0.9797
	B	*0.582*	–	–	–	–	*843.8619*	*0.9962*	B	*0.8796*	–	–	–	–	*907.911*	*0.989*

Tab. B.15: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Elvas, Évora, Faro, Lisboa S2, Lousã, Manteigas, Matosinhos and Mértola, in Summer. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Summer															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Elvas	LN	0.0008	rej.	–	–	rej.	827.9855	0.9808	LN	0.149	not rej.	–	–	rej.	863.498	0.9887
	G	0.0091	rej.	rej.	rej.	–	822.0904	0.9916	G	*0.2686*	*not rej.*	*not rej.*	*not rej.*	–	***859.0912***	*0.9873*
	W	0.0309	not rej.	rej.	rej.	–	831.1439	0.9779	W	0.0301	rej.	rej.	rej.	–	868.3629	0.9775
	B	*0.0692*	–	–	–	–	***821.946***	*0.9948*	B	0.2625	–	–	–	–	863.5418	0.9865
Évora	LN	0.1502	rej.	–	–	rej.	853.1138	0.9628	LN	0	rej.	–	–	rej.	838.6196	0.9329
	G	0.4334	not rej.	rej.	rej.	–	842.8138	0.989	G	0.0003	rej.	rej.	rej.	–	823.8781	0.964
	W	0.743	not rej.	not rej.	not rej.	–	834.6399	0.9944	W	*0.3444*	*not rej.*	*not rej.*	*not rej.*	–	*804.4396*	*0.9921*
	B	*0.9128*	–	–	–	–	***832.9029***	*0.9936*	B	0.2169	–	–	–	–	**900.287**	0.9496
Faro	LN	0.0284	rej.	–	–	rej.	961.4268	0.9855	LN	0	rej.	–	–	rej.	988.8262	0.9427
	G	0.0847	not rej.	not rej.	not rej.	–	**957.3252**	0.9765	G	0	rej.	rej.	rej.	–	974.3847	0.9735
	W	0.0458	rej.	rej.	rej.	–	964.1502	0.964	W	*0.2176*	*not rej.*	*not rej.*	*not rej.*	–	***951.2871***	*0.9884*
	B	*0.0651*	–	–	–	–	*94.2732*	*0.985*	B	0.14	–	–	–	–	953.3478	0.9916
Lisboa S2	LN	0.6894	not rej.	–	–	rej.	772.9513	0.9446	LN	0.2272	not rej.	–	–	rej.	1081.6108	0.984
	G	*0.834*	*not rej.*	*not rej.*	*not rej.*	–	*761.3753*	*0.9955*	G	0.4169	not rej.	not rej.	not rej.	–	**1075.8674**	0.9733
	W	0.0402	rej.	rej.	rej.	–	765.4493	0.9632	W	*0.1395*	*not rej.*	*rej.*	*rej.*	–	*1080.0347*	*0.9755*
	B	0.779	–	–	–	–	**753.4782**	0.9947	B	0.1818	–	–	–	–	1081.1891	0.9885
Lousã	LN	0.0005	rej.	–	–	rej.	569.1262	0.874	LN	*0.4627*	*not rej.*	–	–	*not rej.*	***707.3085***	*0.9881*
	G	0.0001	rej	rej	rej	–	564.9286	0.718	G	0.2975	not rej.	not rej.	not rej.	–	708.2783	0.9657
	W	0	rej.	rej.	rej.	–	640.6355	0.8423	W	0	rej.	rej.	rej.	–	750.2305	0.9422
	B	*0.0256*	–	–	–	–	***529.1777***	*0.9731*	B	0.5033	–	–	–	–	708.1682	0.9888
Manteigas	LN	0.2944	not rej.	–	–	not rej.	975.703	0.9919	LN	0.0114	rej.	–	–	rej.	772.3304	0.9675
	G	*0.5584*	*not rej.*	*not rej.*	*not rej.*	–	***974.5253***	*0.9668*	G	0.0415	not rej.	rej.	rej.	–	768.0777	0.9059
	W	0.1187	not rej.	rej.	rej.	–	996.3343	0.9636	W	0.0029	rej.	rej.	rej.	–	793.8168	0.9747
	B	0.5928	–	–	–	–	978.4867	0.9917	B	*0.1841*	–	–	–	–	***760.9701***	*0.9558*
Matosinhos	LN	0.0089	not rej.	–	–	rej.	986.7592	0.973	LN	0.1528	rej.	–	–	rej.	1146.6484	0.9802
	G	0.0366	not rej.	rej.	rej.	–	**979.7488**	0.9815	G	0.462	not rej.	rej.	rej.	–	**1139.1061**	0.9964
	W	*0.0391*	*not rej.*	*rej.*	*not rej.*	–	*984.9669*	*0.9601*	W	*0.5721*	*not rej.*	*not rej.*	*not rej.*	–	*1139.3215*	*0.976*
	B	0.0204	–	–	–	–	984.3581	0.9873	B	0.4892	–	–	–	–	1140.9324	0.9964
Mértola	LN	0.0685	not rej.	–	–	rej.	739.4772	0.9737	LN	0.0025	rej.	–	–	rej.	844.2724	0.9568
	G	0.1809	not rej.	not rej.	not rej.	–	741.1981	0.7681	G	0.0127	not rej.	rej.	rej.	–	835.9905	0.8899
	W	0	rej.	rej.	rej.	–	806.5929	0.9511	W	0.0033	not rej.	rej.	rej.	–	855.3015	0.9823
	B	*0.2446*	–	–	–	–	***731.716***	*0.9418*	B	*0.0596*	–	–	–	–	***825.0785***	*0.9305*

Tab. B.16: Fitted distributions to the **observed wind** of 8 stations: Elvas, Évora, Faro, Lisboa S2, Lousã, Manteigas, Matosinhos and Mértola. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn			**Winter**			**Spring**			**Summer**		
	$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\alpha}_{El}$	7.6155	(6.3944, 9.2849)	$\hat{\mu}_{El}$	1.5419	(1.4837, 1.6034)	$\hat{\omega}_{El}$	4.8214	(4.4002, 5.3665)	$\hat{k}_{El}$	2.4682	(1.3513, 9.6864)
$\hat{\beta}_{El}$	1.5506	(1.2922, 1.9075)	$\hat{\sigma}_{El}$	0.439	(0.3961, 0.4806)	$\hat{\delta}_{El}$	6.4974	(6.3144, 6.68)	$\hat{c}_{El}$	6.3499	(5.334, 7.6537)
–	–	–	–	–	–	–	–	–	$\hat{\lambda}_{El}$	0.1319	(0.0973, 0.1511)
$\hat{\mu}_{Ev}$	1.794	(1.7426, 1.8489)	$\hat{\alpha}_{Ev}$	5.8318	(4.9482, 7.0465)	$\hat{\alpha}_{Ev}$	14.5814	(12.3716, 17.6842)	$\hat{k}_{Ev}$	4.0715	(1.8982, 77.5979)
$\hat{\sigma}_{Ev}$	0.4154	(0.3783, 0.4523)	$\hat{\beta}_{Ev}$	0.8988	(0.7541, 1.095)	$\hat{\beta}_{Ev}$	2.0627	(1.749, 2.5003)	$\hat{c}_{Ev}$	7.1548	(6.1355, 8.5707)
–	–	–	–	–	–	–	–	–	$\hat{\lambda}_{Ev}$	0.104	(0.0627, 0.1202)
$\hat{\mu}_{F}$	1.8885	(1.8354, 1.9354)	$\hat{\mu}_{F}$	1.9248	(1.8784, 1.9718)	$\hat{\omega}_{F}$	4.1884	(3.8115, 4.6594)	$\hat{k}_{F}$	4.4852	(2.0501, 99.875)
$\hat{\sigma}_{F}$	0.3877	(0.3525, 0.4206)	$\hat{\sigma}_{F}$	0.3586	(0.3247, 0.3903)	$\hat{\delta}_{F}$	8.4835	(8.224, 8.7482)	$\hat{c}_{F}$	5.1635	(4.43, 6.1106)
–	–	–	–	–	–	–	–	–	$\hat{\lambda}_{F}$	0.0948	(0.046, 0.116)
$\hat{\mu}_{Lx}$	1.7405	(1.696, 1.7838)	$\hat{\mu}_{Lx}$	1.7388	(1.6916, 1.7841)	$\hat{\alpha}_{Lx}$	23.132	(19.6769, 27.8753)	$\hat{\alpha}_{Lx}$	29.5701	(24.9888, 35.6366)
$\hat{\sigma}_{Lx}$	0.3429	(0.3104, 0.3723)	$\hat{\sigma}_{Lx}$	0.3499	(0.3164, 0.3811)	$\hat{\beta}_{Lx}$	3.6116	(3.0683, 4.3681)	$\hat{\beta}_{Lx}$	4.5868	(3.881, 5.5459)
$\hat{k}_{Lo}$	0.482	(0.326, 0.7455)	$\hat{\mu}_{Lo}$	1.0693	(1.0025, 1.1356)	$\hat{k}_{Lo}$	0.4983	(0.3273, 0.7717)	$\hat{k}_{Lo}$	0.7678	(0.5019, 1.2926)
$\hat{c}_{Lo}$	5.3831	(4.3111, 6.9831)	$\hat{\sigma}_{Lo}$	0.4723	(0.4229, 0.5178)	$\hat{c}_{Lo}$	8.3624	(6.7326, 10.9487)	$\hat{c}_{Lo}$	8.3885	(6.9112, 10.6094)
$\hat{\lambda}_{Lo}$	0.4224	(0.3649, 0.4712)	–	–	–	$\hat{\lambda}_{Lo}$	0.35	(0.3189, 0.3768)	$\hat{\lambda}_{Lo}$	0.3319	(0.2982, 0.3569)
$\hat{\alpha}_{Man}$	5.5098	(4.6584, 6.7332)	$\hat{\alpha}_{Man}$	4.8803	(4.1291, 5.9762)	$\hat{\alpha}_{Man}$	8.1744	(6.9107, 9.9983)	$\hat{\alpha}_{Man}$	12.4763	(10.4644, 15.256)
$\hat{\beta}_{Man}$	0.5279	(0.4415, 0.649)	$\hat{\beta}_{Man}$	0.4588	(0.3826, 0.5673)	$\hat{\beta}_{Man}$	0.7699	(0.6477, 0.9447)	$\hat{\beta}_{Man}$	1.3164	(1.1005, 1.619)
$\hat{\mu}_{Mat}$	1.8338	(1.7878, 1.8803)	$\hat{\mu}_{Mat}$	1.8112	(1.7693, 1.8541)	$\hat{\alpha}_{Mat}$	11.06	(9.379, 13.3624)	$\hat{\omega}_{Mat}$	3.6153	(3.2886, 4.0104)
$\hat{\sigma}_{Mat}$	0.365	(0.3306, 0.3976)	$\hat{\sigma}_{Mat}$	0.318	(0.2877, 0.3469)	$\hat{\beta}_{Mat}$	1.6032	(1.3589, 1.9389)	$\hat{\delta}_{Mat}$	7.1428	(6.8805, 7.4004)
$\hat{\mu}_{Me}$	1.5671	(1.5168, 1.6192)	$\hat{\mu}_{Me}$	1.5272	(1.4737, 1.5801)	$\hat{k}_{Me}$	4.9907	(2.1326, 112.9676)	$\hat{k}_{Me}$	1.3797	(0.8629, 2.9814)
$\hat{\sigma}_{Me}$	0.3881	(0.3513, 0.423)	$\hat{\sigma}_{Me}$	0.4093	(0.3696, 0.4466)	$\hat{c}_{Me}$	4.5754	(3.9271, 5.4579)	$\hat{c}_{Me}$	8.2838	(6.8961, 10.074)
–	–	–	–	–	–	$\hat{\lambda}_{Me}$	0.1168	(0.0513, 0.1493)	$\hat{\lambda}_{Me}$	0.1592	(0.1377, 0.1731)

Tab. B.17: Fitted distributions to the **simulated wind** of 8 stations: Elvas, Évora, Faro, Lisboa S2, Lousã, Manteigas, Matosinhos and Mértola. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn			**Winter**			**Spring**			**Summer**		
	$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\mu}_{El}$	1.6508	(1.5962,1.7023)	$\hat{\mu}_{El}$	1.7606	(1.7025,1.8188)	$\hat{k}_{El}$	1.745	(1.0306,4.3884)	$\hat{\alpha}_{El}$	23.1336	(19.6991,28.074)
$\hat{\sigma}_{El}$	0.3903	(0.3532,0.4363)	$\hat{\sigma}_{El}$	0.4222	(0.3793,0.4639)	$\hat{c}_{El}$	6.3756	(5.3281,7.7767)	$\hat{\beta}_{El}$	3.1515	(2.6717,3.8293)
–	–	–	–	–	–	$\hat{\lambda}_{El}$	0.1204	(0.0971,0.1355)	–	–	–
$\hat{\mu}_{Ev}$	1.658	(1.6088,1.7082)	$\hat{\mu}_{Ev}$	1.7049	(1.6517,1.7557)	$\hat{k}_{Ev}$	2.092	(1.202,6.5588)	$\hat{\omega}_{Ev}$	6.2013	(5.6476,6.867)
$\hat{\sigma}_{Ev}$	0.3933	(0.3577,0.4274)	$\hat{\sigma}_{Ev}$	0.3957	(0.3585,0.4307)	$\hat{c}_{Ev}$	6.8452	(5.7464,8.2493)	$\hat{\delta}_{Ev}$	7.6519	(7.4835,7.8177)
–	–	–	–	–	–	$\hat{\lambda}_{Ev}$	0.1222	(0.0961,0.1374)	–	–	–
$\hat{\mu}_{F}$	2.0235	(1.9783,2.0705)	$\hat{k}_{F}$	1.1523	(0.7277,2.2188)	$\hat{\omega}_{F}$	5.1421	(4.6653,5.7171)	$\hat{\omega}_{F}$	5.3857	(4.917,5.9918)
$\hat{\sigma}_{F}$	0.3682	(0.3341,0.4017)	$\hat{c}_{F}$	5.4715	(4.5443,6.7562)	$\hat{\delta}_{F}$	9.586	(9.341,9.8315)	$\hat{\delta}_{F}$	8.9833	(8.7632,9.196)
–	–	–	$\hat{\lambda}_{F}$	0.1167	(0.0961,0.1319)	–	–	–	–	–	–
$\hat{k}_{Lx}$	2.2455	(1.2651,8.0963)	$\hat{k}_{Lx}$	2.4047	(1.2863,11.4876)	$\hat{\mu}_{Lx}$	2.1244	(2.0905,2.1575)	$\hat{\omega}_{Lx}$	4.4921	(4.0877,4.9974)
$\hat{c}_{Lx}$	3.998	(3.3629,4.8083)	$\hat{c}_{Lx}$	3.8227	(3.1965,4.6166)	$\hat{\sigma}_{Lx}$	0.2669	(0.2425,0.2895)	$\hat{\delta}_{Lx}$	10.1156	(9.8155,10.4193)
$\hat{\lambda}_{Lx}$	0.1089	(0.0685,0.1342)	$\hat{\lambda}_{Lx}$	0.1042	(0.0588,0.131)	–	–	–	–	–	–
$\hat{\mu}_{Lo}$	1.9757	(1.9275,2.0241)	$\hat{\mu}_{Lo}$	1.9892	(1.9382,2.0407)	$\hat{k}_{Lo}$	0.553	(0.3642,0.8722)	$\hat{\mu}_{Lo}$	2.0049	(1.9856,2.0246)
$\hat{\sigma}_{Lo}$	0.3654	(0.331,0.3975)	$\hat{\sigma}_{Lo}$	0.3666	(0.3286,0.4031)	$\hat{c}_{Lo}$	12.4079	(10.0699,16.1023)	$\hat{\sigma}_{Lo}$	0.1513	(0.1378,0.1646)
–	–	–	–	–	–	$\hat{\lambda}_{Lo}$	0.1379	(0.1295,0.1448)	–	–	–
$\hat{\alpha}_{Man}$	8.0783	(6.7463,9.9292)	$\hat{\alpha}_{Man}$	7.4323	(6.2257,9.1652)	$\hat{k}_{Man}$	0.9467	(0.6146,1.6658)	$\hat{k}_{Man}$	1.8835	(1.0575,6.1457)
$\hat{\beta}_{Man}$	0.9102	(0.7583,1.1285)	$\hat{\beta}_{Man}$	0.8144	(0.676,1.0126)	$\hat{c}_{Man}$	7.9713	(6.5801,9.9569)	$\hat{c}_{Man}$	8.1251	(6.7155,10.0558)
–	–	–	–	–	–	$\hat{\lambda}_{Man}$	0.1091	(0.09701,0.1181)	$\hat{\lambda}_{Man}$	0.1076	(0.0868,0.119)
$\hat{\alpha}_{Mat}$	6.3406	(5.3831,7.7009)	$\hat{\mu}_{Mat}$	2.1369	(2.0905,2.1842)	$\hat{\alpha}_{Mat}$	10.9235	(9.2977,13.0673)	$\hat{\omega}_{Mat}$	3.7789	(3.4473,4.2087)
$\hat{\beta}_{Mat}$	0.7032	(0.5939,0.8636)	$\hat{\sigma}_{Mat}$	0.3609	(0.327,0.3937)	$\hat{\beta}_{Mat}$	1.1598	(0.978,1.3961)	$\hat{\delta}_{Mat}$	10.2104	(9.8406,10.5639)
$\hat{\mu}_{Me}$	1.6948	(1.6417,1.7484)	$\hat{\alpha}_{Me}$	6.6887	(5.6560,8.1139)	$\hat{k}_{Me}$	3.9178	(1.819,79.6683)	$\hat{k}_{Me}$	2.5011	(1.3742,11.542)
$\hat{\sigma}_{Me}$	0.4095	(0.3714,0.4468)	$\hat{\beta}_{Me}$	1.0442	(0.8764,1.2745)	$\hat{c}_{Me}$	5.5986	(4.7579,6.729)	$\hat{c}_{Me}$	7.5912	(6.3733,9.1081)
–	–	–	–	–	–	$\hat{\lambda}_{Me}$	0.0971	(0.0502,0.1168)	$\hat{\lambda}_{Me}$	0.1121	(0.0844,0.125)

Tab. B.18: Copulas selected according to the AIC to fit 8 stations: Elvas, Évora, Faro, Lisboa S2, Lousã, Manteigas, Matosinhos and Mértola. $\hat{\theta}_1$ and $\hat{\theta}_2$ are the copula parameters estimates obtained by MPLE.

Station	Season	Copula	$\hat{\theta}_{MPLE}$				Tail Dependence [1]		p-values		
			$\hat{\theta}_1$	CI (95%)	$\hat{\theta}_2$	CI (95%)	$\hat{\lambda}_L$	$\hat{\lambda}_U$	$S_n^{(B)}$	S_n	$S_n^{(K)}$
Elvas	Autumn	C_α^{gu}	1.8856	(1.6258,2.1454)	–	–	0	0.5557	0.986	1	0.53
	Winter	C_α^{gu}	2.386	(2.0985,2.6735)	–	–	0	0.6629	0.997	0.995	0.5
	Spring	C_α^{f}	4.1116	(3.2163,5.007)	–	–	0	0	0.83	1	0.54
	Summer	C_ρ^{gaus}	0.6237	(0.5378,0.7095)	–	–	0	0	0.92	1	0.67
Évora	Autumn	C_α^{gu}	2.5996	(2.2957,2.9035)	–	–	0	0.6944	0.972	1	0.55
	Winter	C_α^{gu}	2.7014	(2.3392,3.0636)	–	–	0	0.7075	0.943	0.999	0.37
	Spring	C_α^{gu}	2.0865	(1.8061,2.3668)	–	–	0	0.6059	1	1	0.57
	Summer	$C_{\rho\eta}^{t}$	0.6024	(0.5128,0.692)	3.8694	(*)	0.323	0.323	0.969	0.7957	0.32
Faro	Autumn	C_α^{gu}	2.3919	(2.0971,2.6868)	–	–	0	0.6639	0.817	0.975	0.73
	Winter	C_α^{gu}	2.4074	(2.1283,2.6865)	–	–	0	0.6663	0.982	1	0.53
	Spring	C_α^{gu}	1.8201	(1.57,2.0702)	–	–	0	0.5365	0.7	1	0.95
	Summer	C_ρ^{gaus}	0.57	(0.4883,0.6516)	–	–	0	0	0.992	1	0.51
Lisboa S2	Autumn	C_α^{gu}	2.1194	(1.8262,2.4127)	–	–	0	0.6131	0.996	1	0.66
	Winter	C_α^{gu}	2.1036	(1.7683,2.4389)	–	–	0	0.6097	0.997	1	0.7
	Spring	C_ρ^{gaus}	0.5651	(0.4714,0.6587)	–	–	0	0	0.954	1	0.18
	Summer	C_α^{gu}	1.6496	(1.4776,1.8217)	–	–	0	0.4778	0.867	1	0.21
Lousã	Autumn	C_α^{gu}	1.788	(1.5482,2.0279)	–	–	0	0.5265	0.524	1	0.77
	Winter	C_α^{sc}	1.1886	(0.8548,1.5224)	–	–	0	0.5581	0.265	0.165	1
	Spring	C_α^{gu}	1.4637	(1.2928,1.6346)	–	–	0	0.3943	0.939	1	0.49
	Summer	C_α^{gu}	1.191	(1.0759,1.306)	–	–	0	0.2104	0.703	1	0.92
Manteigas	Autumn	C_α^{gu}	1.777	(1.5904,1.9651)	–	–	0	0.5232	0.953	0.99	0.76
	Winter	C_α^{f}	4.9561	(3.9021,6.0101)	–	–	0	0	0.999	1	0.43
	Spring	C_α^{f}	5.7515	(4.5782,6.9246)	–	–	0	0	0.997	1	0.82
	Summer	C_α^{f}	3.7408	(2.8476,4.6341)	–	–	0	0	1	1	0.14
Matosinhos	Autumn	C_α^{gu}	2.4777	(2.2095,2.7459)	–	–	0	0.6772	0.985	0.993	0.64
	Winter	C_α^{gu}	2.312	(1.8167,2.3296)	–	–	0	0.6504	0.86	1	0.48
	Spring	C_α^{gu}	2.0732	(1.8167,2.3296)	–	–	0	0.603	1	0.997	0.71
	Summer	C_α^{f}	5.79	(4.6537,6.9263)	–	–	0	0	0.784	1	0.43
Mértola	Autumn	C_ρ^{gaus}	0.7901	(0.7364,0.8437)	–	–	0	0	0.907	0.979	0.49
	Winter	C_α^{gu}	2.5101	(2.2361,2.784)	–	–	0	0.682	0.872	1	0.42
	Spring	$C_{\rho\eta}^{t}$	0.6552	(0.5798,0.7305)	4.0181	(*)	0.3533	0.3533	0.975	0.0145	0.42
	Summer	$C_{\rho\eta}^{t}$	0.5681	(0.4799,0.6562)	4.6399	(*)	0.2619	0.2619	0.986	0.3551	0.61

(*) At present the asymptotic variance cannot be fully estimated if η is not fixed, thus it is not possible to provide a confidence interval; see the package *copula* manual.
[1] the 0's are theoretical.

Tab. B.19: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Montalegre, Moura, Odemira, Peniche, Portalegre, Proença-a-Nova, Rio Maior and Santarém, in Autumn. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Autumn															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Montalegre	LN	0.7659	not rej.	–	–	not rej.	941.8763	0.9942	LN	0.0488	not rej.	–	–	rej.	1014.6532	0.9844
	G	*0.8625*	*not rej.*	*not rej.*	*not rej.*	–	***940.8654***	*0.9922*	G	*0.1421*	*not rej.*	*not rej.*	*not rej.*	–	***1008.113***	*0.9868*
	W	0.2449	not rej.	rej.	rej.	–	955.6561	0.969	W	0.0577	not rej.	not rej.	not rej.	–	1015.9379	0.9825
	B	0.9297	–	–	–	–	946.4312	0.9883	B	0.1044	–	–	–	–	1010.9946	0.9898
Moura	LN	*0.005*	*not rej.*	–	–	*rej.*	*900.9601*	*0.9419*	LN	*0.1848*	*not rej.*	–	–	*rej.*	***962.0694***	*0.9869*
	G	0.0002	rej.	rej.	rej.	–	896.2256	0.9823	G	0.0126	rej.	rej.	rej.	–	972.2534	0.9841
	W	0	rej.	rej.	rej.	–	924.1541	0.9126	W	0	rej.	rej.	rej.	–	1992.0475	0.9044
	B	0.0032	–	–	–	–	**888.3376**	0.9816	B	0.2473	–	–	–	–	973.8851	0.9557
Odemira	LN	*0.0012*	*rej.*	–	–	*rej.*	***991.7812***	*0.9889*	LN	*0.1648*	*not rej.*	–	–	*rej.*	***1112.0598***	*0.989*
	G	0	rej.	rej.	rej.	–	1001.7819	0.9889	G	0.0496	not rej.	rej.	rej.	–	1120.5432	0.9874
	W	0	rej.	rej.	rej.	–	1036.7649	0.9082	W	0.0001	rej.	rej.	rej.	–	1152.4916	0.9075
	B	0.0008	–	–	–	–	1000.0822	0.9654	B	0.0562	–	–	–	–	1124.9548	0.9697
Peniche	LN	0.2779	not rej.	–	–	not rej.	1127.7927	0.9944	LN	*0.3053*	*not rej.*	–	–	*rej.*	*1057.2332*	*0.9872*
	G	*0.2721*	*not rej.*	*not rej.*	*not rej.*	–	***1126.9433***	*0.9952*	G	0.1963	not rej.	not rej.	not rej.	–	**1055.1818**	0.9797
	W	0.0183	rej.	rej.	rej.	–	1142.3697	0.9637	W	0.0005	rej.	rej.	rej.	–	1073.3241	0.9648
	B	0.166	–	–	–	–	1134.9456	0.9889	B	0.3183	–	–	–	–	1059.3867	0.9729
Portalegre	LN	*0.0005*	*not rej.*	–	–	*not rej.*	***1084.2842***	*0.9899*	LN	*0.6356*	*not rej.*	–	–	*not rej.*	***1087.8741***	*0.9954*
	G	0.0002	rej.	rej.	rej.	–	1084.7654	0.986	G	0.4249	not rej.	not rej.	not rej.	–	1089.3117	0.9892
	W	0	rej.	rej.	rej.	–	1108.894	0.9577	W	0.0024	rej.	rej.	rej.	–	1113.8186	0.9561
	B	0	–	–	–	–	1089.5201	0.9906	B	0.5026	–	–	–	–	1094.5559	0.9795
Proença-a-Nova	LN	*0.3002*	*not rej.*	–	–	*not rej.*	***921.5509***	*0.997*	LN	0.6908	not rej.	–	–	not rej.	997.3721	0.9947
	G	0.1787	not rej.	not rej.	not rej.	–	926.4382	0.9864	G	*0.6693*	*not rej.*	*not rej.*	*not rej.*	–	***996.3526***	*0.9897*
	W	0.011	rej.	rej.	rej.	–	955.3028	0.9425	W	0.0213	rej.	rej.	rej.	–	1021.4172	0.9671
	B	0.1187	–	–	–	–	930.3598	0.9932	B	0.6345	–	–	–	–	999.2426	0.9972
Rio Maior	LN	0	rej.	–	–	rej.	772.9322	0.8117	LN	0.0295	rej.	–	–	rej.	835.024	0.9777
	G	0.0128	rej.	rej.	rej.	–	723.8332	0.9611	G	0.131	not rej.	rej.	rej.	–	830.0771	0.9733
	W	0.2029	not rej.	rej.	not rej.	–	711.8279	0.899	W	0.032	rej.	rej.	rej.	–	845.7071	0.978
	B	*0.1417*	–	–	–	–	***709.717***	*0.9516*	B	*0.2532*	–	–	–	–	***828.6024***	*0.9913*
Santarém	LN	0.2907	rej.	–	–	rej.	922.1726	0.9847	LN	0.0913	rej.	–	–	rej.	1104.6983	0.982
	G	0.0177	rej.	rej.	rej.	–	937.9754	0.9462	G	0.5475	not rej.	not rej.	not rej.	–	**1095.987**	0.9948
	W	0	rej.	rej.	rej.	–	979.1792	0.9136	W	0.6395	not rej.	not rej.	not rej.	–	1102.7047	0.9822
	B	*0.6043*	–	–	–	–	***918.0707***	*0.9919*	B	*0.7434*	–	–	–	–	*1098.5199*	*0.9962*

Tab. B.20: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Montalegre, Moura, Odemira, Peniche, Portalegre, Proença-a-Nova, Rio Maior and Santarém, in Winter. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Winter															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Montalegre	LN	0.428	not rej.	–	–	not rej.	**1006.9688**	0.9937	LN	0.7376	not rej.	–	–	not rej.	1010.1556	0.993
	G	0.1895	not rej.	not rej.	not rej.	–	1009.3513	0.9861	G	*0.9058*	*not rej.*	*not rej.*	*not rej.*	–	***1007.3999***	*0.9926*
	W	0.0037	rej.	rej.	rej.	–	1027.501	0.9594	W	0.5074	not rej.	rej.	rej.	–	1019.3718	0.9675
	B	*0.4622*	–	–	–	–	*1012.0532*	*0.9827*	B	0.8102	–	–	–	–	1015.5219	0.9894
Moura	LN	*0.0052*	*not rej.*	–	–	*not rej.*	***828.7388***	*0.993*	LN	*0.0342*	*rej.*	–	–	*rej.*	***938.4147***	*0.9805*
	G	0.0001	rej.	rej.	rej.	–	836.0844	0.9776	G	0.0023	rej.	rej.	rej.	–	953.17	0.9789
	W	0	rej.	rej.	rej.	–	867.7235	0.9346	W	0	rej.	rej.	rej.	–	989.2888	0.8792
	B	0.0031	–	–	–	–	833.9465	0.9871	B	0.0102	–	–	–	–	945.2728	0.959
Odemira	LN	*0.7911*	*not rej.*	–	–	*not rej.*	***999.2723***	*0.9941*	LN	*0.0804*	*not rej.*	–	–	*not rej.*	*1095.0385*	*0.9879*
	G	0.6564	not rej.	not rej.	not rej.	–	1001.8676	0.9906	G	0.0237	not rej.	rej.	rej.	–	**1094.8201**	0.9747
	W	0.0453	rej.	rej.	rej.	–	1023.4757	0.9443	W	0	rej.	rej.	rej.	–	1113.9167	0.9537
	B	0.3151	–	–	–	–	1010.7974	0.9799	B	0.0523	–	–	–	–	1102.9631	0.9568
Peniche	LN	0.0435	not rej.	–	–	rej.	1162.0358	0.9871	LN	0.8025	not rej.	–	–	not rej.	1077.2704	0.9911
	G	0.4211	not rej.	not rej.	not rej.	–	**1155.5174**	0.9945	G	0.9082	not rej.	not rej.	not rej.	–	**1077.2704**	0.9851
	W	0.6199	not rej.	rej.	rej.	–	1164.3804	0.9783	W	0.1314	rej.	rej.	rej.	–	1095.633	0.9712
	B	*0.5607*	–	–	–	–	*1160.7677*	*0.994*	B	*0.939*	–	–	–	–	*1077.7037*	*0.9873*
Portalegre	LN	0.0426	not rej.	–	–	not rej.	**1071.0543**	0.9924	LN	0.1019	not rej.	–	–	not rej.	1060.1786	0.9896
	G	0.2311	not rej.	not rej.	not rej.	–	1071.2731	0.9721	G	0.2292	not rej.	not rej.	not rej.	–	1061.0591	0.9851
	W	0.0857	not rej.	rej.	rej.	–	1095.9216	0.9595	W	0.0017	rej.	rej.	rej.	–	1096.0251	0.9588
	B	*0.1734*	–	–	–	–	*1074.6344*	*0.9959*	B	*0.3048*	–	–	–	–	***1060.1731***	*0.9895*
Proença-a-Nova	LN	0.2359	not rej.	–	–	not rej.	986.4939	0.9868	LN	0.5428	not rej.	–	–	not rej.	**1005.8542**	0.9883
	G	0.1937	not rej.	not rej.	not rej.	–	987.6004	0.9528	G	0.2973	not rej.	rej.	rej.	–	1010.5303	0.97
	W	0.0025	rej.	rej.	rej.	–	1014.9577	0.9576	W	0.0002	rej.	rej.	rej.	–	1048.5968	0.9361
	B	*0.3331*	–	–	–	–	***986.4307***	*0.9931*	B	*0.6261*	–	–	–	–	*1005.9914*	*0.9899*
Rio Maior	LN	0.0602	rej.	–	–	rej.	740.8244	0.9208	LN	*0.7874*	*not rej.*	–	–	*not rej.*	***846.7201***	*0.9931*
	G	0.3489	not rej.	not rej.	not rej.	–	716.4246	0.9919	G	0.355	not rej.	rej.	rej.	–	849.4678	0.9775
	W	*0.4267*	*not rej.*	*not rej.*	*not rej.*	–	***712.0116***	*0.9646*	W	0.0006	rej.	rej.	rej.	–	871.298	0.944
	B	0.3816	–	–	–	–	713.3922	0.9961	B	0.7037	–	–	–	–	854.3043	0.9657
Santarém	LN	*0.2744*	*not rej.*	–	–	*rej.*	***807.6626***	*0.9851*	LN	*0.1899*	*not rej.*	–	–	*rej.*	***981.2965***	*0.9862*
	G	0.0245	rej.	rej.	rej.	–	822.5557	0.954	G	0.0753	not rej.	rej.	rej.	–	981.5089	0.992
	W	0	rej.	rej.	rej.	–	864.4788	0.8934	W	0.0005	rej.	rej.	rej.	–	999.1339	0.9458
	B	0.4588	–	–	–	–	809.1607	0.9882	B	0.0184	–	–	–	–	991.2032	0.975

Tab. B.21: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Montalegre, Moura, Odemira, Peniche, Portalegre, Proença-a-Nova, Rio Maior and Santarém, in Spring. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Spring															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Montalegre	LN	0.0605	not rej.	–	–	rej.	921.3697	0.9633	LN	*0.115*	*not rej.*	–	–	*not rej.*	*923.2415*	*0.9908*
	G	0.0047	rej.	rej.	rej.	–	921.1063	0.9754	G	0.0883	not rej.	not rej.	rej.	–	**921.9534**	0.9943
	W	0	rej.	rej.	rej.	–	947.6149	0.9278	W	0.0003	rej.	rej.	rej.	–	939.9597	0.9583
	B	*0.1227*	–	–	–	–	***916.0949***	*0.9692*	B	0.0126	–	–	–	–	931.2831	0.988
Moura	LN	0.3268	not rej.	–	–	rej.	894.5318	0.9129	LN	0.0106	rej.	–	–	rej.	900.4272	0.9846
	G	*0.5184*	*not rej.*	*not rej.*	*not rej.*	–	*874.2855*	*0.9927*	G	*0.0329*	*not rej.*	*rej.*	*rej.*	–	***895.8614***	*0.9885*
	W	0.0312	rej.	rej.	rej.	–	883.0568	0.9256	W	0.0078	not rej.	rej.	rej.	–	905.601	0.9694
	B	0.245	–	–	–	–	**871.4667**	0.9932	B	0.0395	–	–	–	–	900.5492	0.9877
Odemira	LN	0.489	not rej.	–	–	rej.	937.9604	0.9791	LN	0.0415	not rej.	–	–	rej.	1040.6095	0.989
	G	*0.684*	*not rej.*	*not rej.*	*not rej.*	–	***930.8189***	*0.9953*	G	*0.0511*	*not rej.*	*rej.*	*not rej.*	–	***1038.1104***	*0.9797*
	W	0.1865	not rej.	rej.	rej.	–	942.4787	0.9774	W	0.0032	not rej.	rej.	rej.	–	1048.359	0.9598
	B	0.5272	–	–	–	–	932.5706	0.997	B	0.0045	–	–	–	–	1047.9879	0.9887
Peniche	LN	0.3883	not rej.	–	–	not rej.	1115.6012	0.9936	LN	0.1244	rej.	–	–	not rej.	1040.6192	0.9924
	G	*0.5888*	*not rej.*	*not rej.*	*not rej.*	–	***1112.467***	*0.9954*	G	*0.2945*	*not rej.*	*not rej.*	*not rej.*	–	***1038.3962***	*0.99*
	W	0.1083	not rej.	rej.	rej.	–	1130.0714	0.9712	W	0.0621	rej.	rej.	rej.	–	1060.2703	0.9678
	B	0.4129	–	–	–	–	1118.5162	0.9973	B	0.3699	–	–	–	–	1042.7715	0.9935
Portalegre	LN	*0.4647*	*not rej.*	–	–	*not rej.*	***895.8062***	*0.9941*	LN	0.7778	not rej.	–	–	not rej.	989.9408	0.9949
	G	0.3578	not rej.	not rej.	not rej.	–	898.1593	0.9825	G	*0.8083*	*not rej.*	*not rej.*	*not rej.*	–	***988.3678***	*0.9902*
	W	0.0007	rej.	rej.	rej.	–	935.0382	0.9448	W	0.0732	rej.	rej.	rej.	–	1011.065	0.9647
	B	0.2576	–	–	–	–	900.8562	0.9956	B	0.7348	–	–	–	–	994.6374	0.9928
Proença-a-Nova	LN	0.0616	rej.	–	–	rej.	898.2614	0.9737	LN	0.2783	not rej.	–	–	not rej.	946.4425	0.9903
	G	0.1971	rej.	not rej.	not rej.	–	890.0055	0.9344	G	*0.51*	*not rej.*	*not rej.*	*not rej.*	–	***942.4989***	*0.9965*
	W	0.0567	not rej.	rej.	rej.	–	904.4492	0.9854	W	0.2697	not rej.	rej.	rej.	–	956.89	0.9796
	B	*0.2457*	–	–	–	–	***887.2392***	*0.959*	B	0.7	–	–	–	–	945.0964	0.9964
Rio Maior	LN	0.0137	rej.	–	–	rej.	725.2168	0.9044	LN	0.5476	not rej.	–	–	rej.	759.0699	0.9712
	G	0.1422	not rej.	rej.	rej.	–	701.6636	0.9801	G	0.7794	not rej.	not rej.	not rej.	–	752.3166	0.9965
	W	*0.3678*	*not rej.*	*not rej.*	*not rej.*	–	***684.143***	*0.9679*	W	0.7016	not rej.	rej.	rej.	–	752.989	0.9882
	B	0.325	–	–	–	–	685.867	0.9919	B	*0.9369*	–	–	–	–	***748.2112***	*0.9975*
Santarém	LN	*0.5802*	*not rej.*	–	–	*rej.*	*808.8967*	*0.9872*	LN	0	rej.	–	–	rej.	1043.2909	0.9439
	G	0.315	not rej.	rej.	rej.	–	818.0645	0.9413	G	0.0001	rej.	rej.	rej.	–	1028.8539	0.9822
	W	0	rej.	rej.	rej.	–	872.9991	0.9134	W	*0.0399*	*not rej.*	*not rej.*	*not rej.*	–	***1007.8504***	*0.9945*
	B	0.5187	–	–	–	–	**807.5111**	0.9927	B	0.03	–	–	–	–	1009.0362	0-9926

Tab. B.22: Goodness-of-fit tests' results for the marginal distributions fitted to the 8 stations: Montalegre, Moura, Odemira, Peniche, Portalegre, Proença-a-Nova, Rio Maior and Santarém, in Summer. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Summer															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Montalegre	LN	*0.8122*	*not rej.*	–	–	*rej.*	*834.278*	*0.9818*	LN	0.1208	not rej.	–	–	not rej.	869.6468	0.9919
	G	0.7534	not rej.	not rej.	not rej.	–	**830.7321**	0.9801	G	*0.2097*	*not rej.*	*not rej.*	*not rej.*	–	***867.1816***	*0.9966*
	W	0.0119	not rej.	rej.	rej.	–	853.6491	0.9665	W	0.0267	rej.	rej.	rej.	–	885.3013	0.9717
	B	0.6657	–	–	–	–	831.2463	0.9975	B	0.2015	–	–	–	–	871.7128	0.9936
Moura	LN	0.0711	not rej.	–	–	not rej.	736.3444	0.9958	LN	0	rej.	–	–	rej.	854.6746	0.964
	G	*0.1211*	*not rej.*	*not rej.*	*not rej.*	–	***736.0548***	*0.9896*	G	0	rej.	rej.	rej.	–	845.6481	0.9486
	W	0.0044	rej.	rej.	rej.	–	762.9101	0.959	W	0.0062	not rej.	rej.	not rej.	–	847.2108	0.9826
	B	0.0625	–	–	–	–	742.0509	0.9953	B	*0.002*	–	–	–	–	***839.2498***	*0.9413*
Odemira	LN	*0.0174*	*rej.*	–	–	*rej.*	***879.4475***	*0.9819*	LN	0.0343	not rej.	–	–	rej.	1022.181	0.9904
	G	0.0028	rej.	rej.	rej.	–	883.9145	0.9819	G	*0.0306*	*not rej.*	*rej.*	*rej.*	–	***1020.941***	*0.9816*
	W	0	rej.	rej.	rej.	–	912.868	0.9041	W	0.0002	rej.	rej.	rej.	–	1035.537	0.9533
	B	0.0077	–	–	–	–	896.4242	0.9349	B	0.0022	–	–	–	–	1032.6772	0.9814
Peniche	LN	0.2435	rej.	–	–	rej.	1076.139	0.9202	LN	0.0451	not rej.	–	–	rej.	1037.8818	0.9873
	G	0.693	not rej.	not rej.	not rej.	–	1052.4413	0.9848	G	*0.0736*	*not rej.*	*rej.*	*rej.*	–	***1035.6633***	*0.9774*
	W	0.4127	not rej.	rej.	not rej.	–	1051.646	0.9555	W	0.012	rej.	rej.	rej.	–	1046.6288	0.9547
	B	*0.629*	–	–	–	–	***1045.4889***	*0.9888*	B	0.0126	–	–	–	–	1046.3342	0.9866
Portalegre	LN	*0.7207*	*not rej.*	–	–	*not rej.*	***878.7842***	*0.9971*	LN	0.0029	rej.	–	–	rej.	900.7467	0.9631
	G	0.3891	not rej.	not rej.	not rej.	–	883.1539	0.985	G	0.024	rej.	rej.	rej.	–	890.331	0.9884
	W	0.0001	rej.	rej.	rej.	–	923.3634	0.9342	W	0.3429	not rej.	not rej.	not rej.	–	878.901	0.9955
	B	0.674	–	–	–	–	885.9331	0.9919	B	*0.3431*	–	–	–	–	***878.4021***	*0.9952*
Proença-a-Nova	LN	0.5276	not rej.	–	–	rej.	773.5262	0.9477	LN	0.8816	not rej.	–	–	not rej.	896.2541	0.9945
	G	0.6729	not rej.	not rej.	not rej.	–	762.4191	0.9809	G	*0.8729*	*not rej.*	*not rej.*	*not rej.*	–	***895.8224***	*0.9824*
	W	0.0302	rej.	rej.	rej.	–	768.9749	0.9716	W	0.0701	not rej.	rej.	rej.	–	916.029	0.9523
	B	*0.5103*	–	–	–	–	***753.1641***	*0.9893*	B	0.6119	–	–	–	–	906.6987	0.9753
Rio Maior	LN	0.0164	rej.	–	–	rej.	715.9259	0.8997	LN	0.5977	not rej.	–	–	not rej.	681.6412	0.9885
	G	0.1719	rej.	rej.	rej.	–	694.0395	0.9922	G	*0.6434*	*not rej.*	*not rej.*	*not rej.*	–	***680.0581***	*0.9875*
	W	0.1602	rej.	rej.	rej.	–	677.3996	0.9704	W	0.0495	not rej.	rej.	rej.	–	700.0094	0.9632
	B	*0.4236*	–	–	–	–	***676.4511***	*0.9915*	B	0.3823	–	–	–	–	684.4898	0.9922
Santarém	LN	0.0013	rej.	–	–	rej.	675.3897	0.9357	LN	0.0131	not rej.	–	–	not rej.	925.5515	0.9901
	G	0.0014	rej.	rej.	rej.	–	670.1985	0.935	G	*0.0167*	*not rej.*	*not rej.*	*not rej.*	–	***922.487***	*0.9848*
	W	0	rej.	rej.	rej.	–	709.685	0.9257	W	0.0004	not rej.	rej.	rej.	–	939.3699	0.9754
	B	*0.0219*	–	–	–	–	***652.6085***	*0.9784*	B	0.0027	–	–	–	–	925.5868	0.9913

Tab. B.23: Fitted distributions to the **observed wind** of 8 stations: Montalegre, Moura, Odemira, Peniche, Portalegre, Proença-a-Nova, Rio Maior and Santarém. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn			**Winter**			**Spring**			**Summer**		
	$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\alpha}_{Mon}$	5.0152	(4.1892, 6.1452)	$\hat{k}_{Mon}$	1.3303	(0.8038, 2.9042)	$\hat{k}_{Mon}$	0.8031	(0.5191, 1.3637)	$\hat{\mu}_{Mon}$	1.6543	(1.6134, 1.6958)
$\hat{\beta}_{Mon}$	0.957	(0.7888, 1.1914)	$\hat{c}_{Mon}$	3.2765	(2.7145, 4.0623)	$\hat{c}_{Mon}$	5.4713	(4.4791, 6.9385)	$\hat{\sigma}_{Mon}$	0.3134	(0.2844, 0.3432)
–	–	–	$\hat{\lambda}_{Mon}$	0.1759	(0.1208, 0.2197)	$\hat{\lambda}_{Mon}$	0.197	(0.1674, 0.2214)	–	–	–
$\hat{\mu}_{Mou}$	1.5057	(1.4566, 1.5537)	$\hat{\mu}_{Mou}$	1.4637	(1.4151, 1.5136)	$\hat{\alpha}_{Mou}$	9.6577	(8.1289, 11.7042)	$\hat{\alpha}_{Mou}$	20.4526	(17.4097, 24.7473)
$\hat{\sigma}_{Mou}$	0.3773	(0.3436, 0.4116)	$\hat{\sigma}_{Mou}$	0.3712	(0.3351, 0.4049)	$\hat{\beta}_{Mou}$	1.792	(1.5002, 2.1947)	$\hat{\beta}_{Mou}$	3.8723	(3.2882, 4.696)
$\hat{\mu}_{O}$	1.7795	(1.7366, 1.8239)	$\hat{\mu}_{O}$	1.8031	(1.7588, 1.8481)	$\hat{\alpha}_{O}$	15.9704	(13.5979, 19.3463)	$\hat{\mu}_{O}$	1.8905	(1.8599, 1.9193)
$\hat{\sigma}_{O}$	0.3432	(0.3104, 0.3739)	$\hat{\sigma}_{O}$	0.3571	(0.323, 0.3895)	$\hat{\beta}_{O}$	2.3072	(1.9601, 2.809)	$\hat{\sigma}_{O}$	0.2372	(0.2151, 0.2582)
$\hat{\alpha}_{Pe}$	6.5586	(5.5457, 8.0423)	$\hat{k}_{Pe}$	2.9436	(1.5329, 27.3519)	$\hat{\alpha}_{Pe}$	10.6017	(8.9915, 12.7309)	$\hat{k}_{Pe}$	3.3502	(1.6572, 45.5843)
$\hat{\beta}_{Pe}$	0.7732	(0.6494, 0.9555)	$\hat{c}_{Pe}$	3.6639	(3.0902, 4.3804)	$\hat{\beta}_{Pe}$	1.3066	(1.103, 1.5821)	$\hat{c}_{Pe}$	4.1716	(3.5445, 4.9961)
–	–	–	$\hat{\lambda}_{Pe}$	0.0826	(0.0374, 0.1061)	–	–	–	$\hat{\lambda}_{Pe}$	0.1028	(0.0468, 0.1298)
$\hat{\mu}_{Po}$	1.83	(1.7833, 1.8788)	$\hat{k}_{Po}$	1.7343	(1.0343, 4.4499)	$\hat{\mu}_{Po}$	1.8482	(1.8186, 1.8786)	$\hat{\mu}_{Po}$	1.8632	(1.8344, 1.8931)
$\hat{\sigma}_{Po}$	0.3791	(0.3451, 0.4126)	$\hat{c}_{Po}$	4.0981	(3.4225, 5.0277)	$\hat{\sigma}_{Po}$	0.2406	(0.2178, 0.2609)	$\hat{\sigma}_{Po}$	0.2341	(0.2126, 0.2542)
–	–	–	$\hat{\lambda}_{Po}$	0.1264	(0.0893, 0.1524)	–	–	–	–	–	–
$\hat{\mu}_{PN}$	1.6038	(1.5571, 1.6496)	$\hat{k}_{PN}$	1.3033	(0.81, 2.6926)	$\hat{k}_{PN}$	2.7332	(1.4646, 16.5723)	$\hat{k}_{PN}$	2.1924	(1.2067, 7.4286)
$\hat{\sigma}_{PN}$	0.3577	(0.325, 0.3903)	$\hat{c}_{PN}$	4.0587	(3.3719, 5.0013)	$\hat{c}_{PN}$	5.568	(4.707, 6.6818)	c_{PN}	7.7052	(6.5094, 9.3798)
–	–	–	$\hat{\lambda}_{PN}$	0.172	(0.1302, 0.204)	$\hat{\lambda}_{PN}$	0.1282	(0.082, 0.15)	$\hat{\lambda}_{PN}$	0.1353	(0.1075, 0.1508)
$\hat{k}_{RM}$	5.5864	(2.1286, 131.9407)	$\hat{\omega}_{RM}$	2.5852	(2.3327, 2.9093)	$\hat{\omega}_{RM}$	4.161	(3.77, 4.6684)	$\hat{k}_{RM}$	4.5668	(1.8922, 115.9896)
$\hat{c}_{RM}$	2.9024	(2.489, 3.5547)	$\hat{\delta}_{RM}$	4.2968	(4.045, 4.5418)	$\hat{\delta}_{RM}$	5.3014	(5.1169, 5.4854)	$\hat{c}_{RM}$	5.0697	(4.3049, 6.1377)
$\hat{\lambda}_{RM}$	0.1313	(0.0364, 0.2034)	–	–	–	–	–	–	$\hat{\lambda}_{RM}$	0.1307	(0.0608, 0.165)
$\hat{k}_{S}$	0.6502	(0.4191, 1.0513)	$\hat{\mu}_{S}$	1.4894	(1.44, 1.5386)	$\hat{\mu}_{S}$	1.6525	(1.6198, 1.6837)	$\hat{k}_{S}$	1.0266	(0.6659, 1.8393)
$\hat{c}_{S}$	5.2043	(4.2769, 6.6819)	$\hat{\sigma}_{S}$	0.36	(0.3254, 0.3929)	$\hat{\sigma}_{S}$	0.2515	(0.2288, 0.2729)	$\hat{c}_{S}$	10.2152	(8.4912, 12.649)
$\hat{\lambda}_{S}$	0.2642	(0.2248, 0.2988)	–	–	–	–	–	–	$\hat{\lambda}_{S}$	0.1857	(0.1692, 0.1972)

Tab. B.24: Fitted distributions to the **simulated wind** of 8 stations: Montalegre, Moura, Odemira, Peniche, Portalegre, Proença-a-Nova, Rio Maior and Santarém. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn			Winter			Spring			Summer		
	$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\alpha}_{Mon}$	10.6292	(8.9323,13.0217)	$\hat{\alpha}_{Mon}$	11.3123	(9.5515,13.8372)	$\hat{\mu}_{Mon}$	2.1708	(2.1411,2.2016)	$\hat{\alpha}_{Mon}$	23.5025	(19.8495,28.8834)
$\hat{\beta}_{Mon}$	1.2357	(1.0361,1.518)	$\hat{\beta}_{Mon}$	1.3361	(1.1242,1.6426)	$\hat{\sigma}_{Mon}$	0.2251	(0.2029,0.2458)	$\hat{\beta}_{Mon}$	2.7025	(2.2713,3.3296)
$\hat{\mu}_{Mou}$	1.6424	(1.5945,1.6907)	$\hat{\mu}_{Mou}$	1.6785	(1.6283,1.7316)	$\hat{\alpha}_{Mou}$	14.0959	(11.8368,17.124)	$\hat{k}_{Mou}$	4.1442	(1.9297,88.062)
$\hat{\sigma}_{Mou}$	0.3758	(0.341,0.4093)	$\hat{\sigma}_{Mou}$	0.3851	(0.3485,0.42)	$\hat{\beta}_{Mou}$	2.0872	(1.7416,2.5469)	$\hat{c}_{Mou}$	5.7416	(4.9195,6.8645)
–	–	–	–	–	–	–	–	–	$\hat{\lambda}_{Mou}$	0.1166	(0.061,0.1395)
$\hat{\mu}_{O}$	2.0454	(2.003,2.09)	$\hat{\mu}_{O}$	2.1083	(2.0641,2.1502)	$\hat{\alpha}_{O}$	20.216	(17.0754,24.493)	$\hat{\alpha}_{O}$	19.7276	(16.6674,24.005)
$\hat{\sigma}_{O}$	0.3409	(0.3085,0.3714)	$\hat{\sigma}_{O}$	0.325	(0.2943,0.3536)	$\hat{\beta}_{O}$	2.0813	(1.7515,2.5328)	$\hat{\beta}_{O}$	2.054	(1.7335,2.5052)
$\hat{\mu}_{Pe}$	2.0778	(2.0349,2.1215)	$\hat{k}_{Pe}$	1.4384	(0.9016,3.1491)	$\hat{\alpha}_{Pe}$	17.6912	(15.0095,21.4309)	$\hat{\alpha}_{Pe}$	15.9294	(13.5393,19.2748)
$\hat{\sigma}_{Pe}$	0.3355	(0.3024,0.3665)	$\hat{c}_{Pe}$	5.4014	(4.4907,6.5857)	$\hat{\beta}_{Pe}$	1.9946	(1.6884,2.4213)	$\hat{\beta}_{Pe}$	1.8653	(1.5821,2.2709)
–	–	–	$\hat{\lambda}_{Pe}$	0.1057	(0.0846,0.1203)	–	–	–	–	–	–
$\hat{\mu}_{Po}$	1.9658	(1.9225,2.0091)	$\hat{k}_{Po}$	1.3481	(0.8491,2.7709)	$\hat{\alpha}_{Po}$	19.7387	(16.6846,23.984)	$\hat{k}_{Po}$	5.9203	(2.38,135.1835)
$\hat{\sigma}_{Po}$	0.3335	(0.3035,0.3624)	$\hat{c}_{Po}$	5.2872	(4.4048,6.4703)	$\hat{\beta}_{Po}$	2.3617	(1.9922,2.8734)	$\hat{c}_{Po}$	6.7818	(5.8826,8.0647)
–	–	–	$\hat{\lambda}_{Po}$	0.1209	(0.0977,0.138)	–	–	–	$\hat{\lambda}_{Po}$	0.0894	(0.0516,0.1064)
$\hat{\alpha}_{PN}$	12.4723	(10.5441,15.0732)	$\hat{k}_{PN}$	0.8655	(0.5614,1.49)	$\hat{\alpha}_{PN}$	30.5405	(25.8473,36.9081)	$\hat{\alpha}_{PN}$	34.7404	(29.3479,42.3093)
$\hat{\beta}_{PN}$	1.6424	(1.3866,1.991)	$\hat{c}_{PN}$	6.5488	(5.4085,8.1889)	$\hat{\beta}_{PN}$	3.197	(2.7015,3.8739)	$\hat{\beta}_{PN}$	3.5588	(3.002,4.3411)
–	–	–	$\hat{\lambda}_{PN}$	0.1348	(0.1178,0.1484)	–	–	–	–	–	–
$\hat{k}_{RM}$	1.7336	(0.9648,5.5631)	$\hat{\mu}_{RM}$	1.9454	(1.8999,1.99)	$\hat{k}_{RM}$	2.792	(1.4169,24.771)	$\hat{\alpha}_{RM}$	43.5946	(36.1988,54.0013)
$\hat{c}_{RM}$	4.8095	(3.928,5.9863)	$\hat{\sigma}_{RM}$	0.314	(0.282,0.3444)	$\hat{c}_{RM}$	7.8957	(6.6111,9.6232)	$\hat{\beta}_{RM}$	4.8264	(4.0142,5.9986)
$\hat{\lambda}_{RM}$	0.123	(0.0861,0.1464)	–	–	–	$\hat{\lambda}_{RM}$	0.0973	(0.0668,0.1096)	–	–	–
$\hat{k}_{S}$	3.4366	(1.6687,53.3702)	$\hat{\mu}_{S}$	1.9413	(1.8966,1.9887)	$\hat{\omega}_{S}$	5.7343	(5.2114,6.3369)	$\hat{\alpha}_{S}$	40.2176	(34.16,49.007)
$\hat{c}_{S}$	3.4613	(2.9263,4.1675)	$\hat{\sigma}_{S}$	0.3444	(0.3109,0.3756)	$\hat{\delta}_{S}$	10.8669	(10.6082,11.1105)	$\hat{\beta}_{S}$	3.7743	(3.2121,4.5967)
$\hat{\lambda}_{S}$	0.0893	(0.0327,0.1181)	–	–	–	–	–	–	–	–	–

Tab. B.25: Copulas selected according to the AIC to fit 8 stations: Montalegre, Moura, Odemira, Peniche, Portalegre, Proença-a-Nova, Rio Maior and Santarém. $\hat{\theta}_1$ and $\hat{\theta}_2$ are the copula parameters estimates obtained by MPLE.

Station	Season	Copula	$\hat{\theta}_{MPLE}$				Tail Dependence [1]		p-values		
			$\hat{\theta}_1$	CI (95%)	$\hat{\theta}_2$	CI (95%)	$\hat{\lambda}_L$	$\hat{\lambda}_U$	$S_n^{(B)}$	S_n	$S_n^{(K)}$
Montalegre	Autumn	C_ρ^{gaus}	0.6323	(0.5502, 0.7144)	–	–	0	0	0.888	1	0.63
	Winter	C_ρ^{gaus}	0.7276	(0.6805, 0.7747)	–	–	0	0	0.87	0.999	0.07
	Spring	C_α^{gu}	1.6002	(1.4119, 1.7885)	–	–	0	0.4579	0.979	0.999	0.18
	Summer	$C_{\rho\eta}^{t}$	0.5081	(0.387, 0.6292)	3.9306	(*)	0.2613	0.2613	0.74	0.2562	0.51
Moura	Autumn	C_α^{sc}	2.1348	(1.7075, 2.5621)	–	–	0	0.7228	0.098	0.021	0.98
	Winter	C_α^{gu}	2.2943	(2.0715, 2.5171)	–	–	0	0.6473	0.635	0.999	0.58
	Spring	C_α^{gu}	2.7892	(1.558, 1.9637)	–	–	0	0.7179	0.629	0.999	0.52
	Summer	$C_{\rho\eta}^{t}$	0.5283	(0.432, 0.6246)	9.5822	(*)	0.0992	0.0992	0.967	0.1124	0.44
Odemira	Autumn	$C_{\rho\eta}^{t}$	0.7423	(0.68, 0.8045)	3.4331	(*)	0.4593	0.4593	0.926	0.4491	0.37
	Winter	C_α^{gu}	2.1277	(1.8774, 2.378)	–	–	0	0.6149	0.999	1	0.78
	Spring	C_α^{f}	4.9584	(3.9303, 5.9865)	–	–	0	0	0.675	0.998	0.44
	Summer	C_α^{gu}	1.7407	(1.591, 1.8904)	–	–	0	0.5109	0.831	0.998	0.69
Peniche	Autumn	C_α^{gu}	2.2324	(1.9773, 2.4875)	–	–	0	0.6359	0.88	0.999	0.67
	Winter	C_α^{gu}	2.1346	(1.8724, 2.3969)	–	–	0	0.6164	0.999	1	0.67
	Spring	$C_{\rho\eta}^{t}$	0.736	(0.6707, 0.8014)	5.8057	(*)	0.3438	0.3438	0.999	0.7108	0.85
	Summer	C_α^{f}	7.3453	(6.171, 8.5196)	–	–	0	0	0.988	1	0.41
Portalegre	Autumn	C_α^{gu}	2.064	(1.7934, 2.3341)	–	–	0	0.6009	1	1	0.45
	Winter	C_α^{gu}	2.1216	(1.8408, 2.4024)	–	–	0	0.6136	0.973	0.967	0.64
	Spring	C_ρ^{gaus}	0.6122	(0.5351, 0.6894)	–	–	0	0	0.997	1	0.06
	Summer	C_α^{sg}	1.5042	(1.3405, 1.668)	–	–	0.4147	0	0.996	1	0.98
Proença-a-Nova	Autumn	C_α^{gu}	2.0269	(1.787, 2.2668)	–	–	0	0.5923	0.998	0.998	0.59
	Winter	C_α^{gu}	2.094	(1.8296, 2.3583)	–	–	0	0.6076	0.994	1	0.73
	Spring	$C_{\rho\eta}^{t}$	0.5382	(0.441, 0.6353)	7.1506	(*)	0.1557	0.1557	0.916	0.3472	0.34
	Summer	C_α^{f}	4.3038	(3.3564, 5.2514)	–	–	0	0	0.975	1	0.54
Rio Maior	Autumn	C_ρ^{gaus}	0.6775	(0.5994, 0.7556)	–	–	0	0	0.991	1	0.6
	Winter	C_ρ^{gaus}	0.7175	(0.6501, 0.7848)	–	–	0	0	0.949	1	0.71
	Spring	C_ρ^{gaus}	0.6497	(0.5706, 0.7288)	–	–	0	0	0.994	1	0.73
	Summer	C_ρ^{gaus}	0.6841	(0.6242, 0.744)	–	–	0	0	0.829	0.981	0.14
Santarém	Autumn	C_α^{gu}	2.1691	(1.8907, 2.4475)	–	–	0	0.6235	0.983	1	0.63
	Winter	C_α^{gu}	2.2631	(1.9665, 2.5298)	–	–	0	0.6416	0.698	0.995	0.78
	Spring	C_α^{sg}	1.4988	(1.3404, 1.6572)	–	–	0.412	0	0.632	0.776	0.57
	Summer	$C_{\rho\eta}^{t}$	0.5165	(0.4175, 0.6155)	2.8025	(*)	0.3356	0.3356	0.558	0.1843	0.52

(*) At present the asymptotic variance cannot be fully estimated if η is not fixed, thus it is not possible to provide a confidence interval; see the package *copula* manual.
[1] the 0's are theoretical.

Tab. B.26: Goodness-of-fit tests' results for the marginal distributions fitted to the 6 stations: Santiago do Cacém, Setúbal, Tomar, Torres Vedras, Vila Real and Viseu, in Autumn. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Autumn															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Santiago do Cacém	LN	0.0787	rej.	–	–	rej.	1042.1523	0.8726	LN	0.0286	rej.	–	–	rej.	1069.107	0.9875
	G	0.5194	not rej.	not rej.	not rej.	–	1002.0457	0.989	G	*0.1496*	*not rej.*	*not rej.*	*not rej.*	–	***1063.5836***	*0.9923*
	W	0.1723	not rej.	rej.	rej.	–	1003.558	0.9159	W	0.0425	not rej.	rej.	not rej.	–	1072.663	0.9708
	B	*0.5792*	–	–	–	–	***997.06***	*0.9928*	B	0.1228	–	–	–	–	1070.2407	0.9909
Setúbal	LN	0.1024	rej.	–	–	not rej.	770.4043	0.9888	LN	0.2529	not rej.	–	–	not rej.	**1131.6226**	0.9961
	G	*0.2926*	*not rej.*	*not rej.*	*not rej.*	–	***764.7406***	*0.9877*	G	*0.4842*	*not rej.*	*not rej.*	*not rej.*	–	*1132.5655*	*0.9873*
	W	0.0777	not rej.	rej.	rej.	–	777.9018	0.9789	W	0.052	rej.	rej.	rej.	–	1136.4295	0.9613
	B	0.2383	–	–	–	–	769.1086	0.9927	B	0.4376	–	–	–	–	1136.4295	0.9971
Tomar	LN	0	rej.	–	–	rej.	894.5802	0.953	LN	0.0099	rej.	–	–	not rej.	986.1583	0.9884
	G	0.008	rej.	rej.	rej.	–	876.0842	0.9839	G	*0.0456*	*not rej.*	*not rej.*	*not rej.*	–	***981.7839***	*0.9869*
	W	*0.0121*	*not rej.*	*not rej.*	*not rej.*	–	***871.2139***	*0.9889*	W	0.0093	rej.	rej.	rej.	–	995.9818	0.9783
	B	0.0087	–	–	–	–	871.914	0.98	B	0.0452	–	–	–	–	984.9666	0.9925
Torres Vedras	LN	0.349	not rej.	–	–	not rej.	1039.296	0.989	LN	0.8512	not rej.	–	–	not rej.	**1082.5836**	0.9976
	G	*0.2785*	*not rej.*	*not rej.*	*not rej.*	–	***1037.0278***	*0.9855*	G	*0.9278*	*not rej.*	*not rej.*	*not rej.*	–	*1083.509*	*0.9869*
	W	0.0071	rej.	rej.	rej.	–	1056.5966	0.9683	W	0.1336	not rej.	rej.	rej.	–	1111.6912	0.9601
	B	0.1691	–	–	–	–	1041.4997	0.9924	B	0.8693	–	–	–	–	1087.243	0.9982
Vila Real	LN	0.1833	not rej.	–	–	rej.	**909.4305**	0.9917	LN	0.2354	not rej.	–	–	not rej.	1072.698	0.9936
	G	*0.2122*	*not rej.*	*rej.*	*rej.*	–	*915.8463*	*0.9949*	G	0.3458	not rej.	not rej.	not rej.	–	**1070.9903**	0.9871
	W	0.0352	rej.	rej.	rej.	–	938.7284	0.9279	W	0.0177	rej.	rej.	rej.	–	1093.5382	0.9686
	B	0.0949	–	–	–	–	924.0319	0.9879	B	*0.3736*	–	–	–	–	*1075.1816*	*0.9936*
Viseu	LN	0.7803	not rej.	–	–	not rej.	1111.0406	0.9908	LN	0.1747	not rej.	–	–	rej.	1002.4484	0.9838
	G	*0.7668*	*not rej.*	*not rej.*	*not rej.*	–	***1107.5683***	*0.9919*	G	*0.2338*	*not rej.*	*not rej.*	*not rej.*	–	***997.7229***	*0.991*
	W	0.1087	not rej.	rej.	rej.	–	1122.6189	0.9707	W	0.0061	not rej.	rej.	rej.	–	1017.3181	0.9697
	B	0.3368	–	–	–	–	1115.0876	0.9898	B	0.1397	–	–	–	–	1000.8306	0.9963

Tab. B.27: Goodness-of-fit tests' results for the marginal distributions fitted to the 6 stations: Santiago do Cacém, Setúbal, Tomar, Torres Vedras, Vila Real and Viseu, in Winter. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Winter															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Santiago do Cacém	LN	0.3708	not rej.	–	–	rej.	991.2634	0.9793	LN	0.0771	not rej.	–	–	not rej.	1042.5745	0.9908
	G	*0.8574*	*not rej.*	*not rej.*	*not rej.*	–	***980.4885***	*0.9838*	G	*0.3648*	*not rej.*	*not rej.*	*not rej.*	–	***1041.4526***	*0.9932*
	W	0.6854	not rej.	not rej.	not rej.	–	984.978	0.9833	W	0.2191	not rej.	rej.	rej.	–	1057.7347	0.9577
	B	0.7989	–	–	–	–	985.1208	0.9849	B	0.2848	–	–	–	–	1048.09	0.9965
Setúbal	LN	0.0086	not rej.	–	–	rej.	789.6095	0.9648	LN	*0.2001*	*not rej.*	–	–	*not rej.*	***1089.7608***	*0.9961*
	G	*0.0106*	ırej.	ırej.	ırej.	–	*822.01*	*0.4128*	G	0.0729	not rej.	not rej.	not rej-	–	1092.718	0.9925
	W	0	rej.	rej.	rej.	–	899.7594	0.9065	W	0.0004	rej.	rej.	rej.	–	1115.0263	0.9491
	B	0.0033	–	–	–	–	788.1982	0.9079	B	0.0297	–	–	–	–	1099.7013	0.9883
Tomar	LN	0.0043	rej.	–	–	rej.	896.8376	0.9691	LN	0.4985	not rej.	–	–	not rej.	991.8265	0.9922
	G	0.1896	not rej.	not rej.	not rej.	–	882.0119	0.9874	G	*0.6779*	*not rej.*	*not rej.*	*not rej.*	–	***988.7415***	*0.9948*
	W	*0.6455*	*not rej.*	*not rej.*	*not rej.*	–	***879.2383***	*0.9921*	W	0.2215	not rej.	rej.	not rej.	–	1001.1983	0.9717
	B	0.4107	–	–	–	–	881.9703	0.9947	B	0.4032	–	–	–	–	995.1349	0.9962
Torres Vedras	LN	0.9129	not rej.	–	–	rej.	1011.7369	0.9854	LN	0.0734	not rej.	–	–	not rej.	**1072.4893**	0.9958
	G	*0.8927*	*not rej.*	*not rej.*	*not rej.*	–	***1005.359***	*0.9894*	G	*0.1729*	*not rej.*	*not rej.*	*not rej.*	–	*1072.883*	*0.9952*
	W	0.0758	not rej.	rej.	rej.	–	1016.4517	0.9789	W	0.0175	rej.	rej.	rej.	–	1095.602	0.9546
	B	0.7525	–	–	–	–	1010.9075	0.9847	B	0.1685	–	–	–	–	1079.7058	0.9931
Vila Real	LN	*0.1585*	*not rej.*	–	–	*not rej.*	***944.201***	*0.9905*	LN	*0.3311*	*not rej.*	–	–	*not rej.*	***1074.0803***	*0.9939*
	G	0.006	rej.	rej.	rej.	–	956.7844	0.9872	G	0.2101	not rej.	not rej.	not rej.	–	1080.5521	0.9249
	W	0	rej.	rej.	rej.	–	983.2989	0.914	W	0.0002	rej.	rej.	rej.	–	1120.0514	0.94
	B	0.0427	–	–	–	–	958.1745	0.9785	B	0.175	–	–	–	–	1080.677	0.9908
Viseu	LN	0.5243	not rej.	–	–	rej.	1042.856	0.9915	LN	0.6308	not rej.	–	–	rej.	985.6128	0.9861
	G	*0.5411*	*not rej.*	*not rej.*	*not rej.*	–	***1041.2759***	*0.9866*	G	0.9693	not rej.	not rej.	not rej.	–	**979.0327**	0.991
	W	0.0427	rej.	rej.	rej.	–	1054.328	0.9584	W	0.7816	not rej.	rej.	rej.	–	991.5421	0.9799
	B	0.2145	–	–	–	–	1051.783	0.9823	B	*0.9812*	–	–	–	–	*981.9716*	*0.9968*

Tab. B.28: Goodness-of-fit tests' results for the marginal distributions fitted to the 6 stations: Santiago do Cacém, Setúbal, Tomar, Torres Vedras, Vila Real and Viseu, in Spring. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Spring															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Santiago do Cacém	LN	0.0033	rej.	–	–	rej.	819.4967	0.9274	LN	0.0541	rej.	–	–	not rej.	816.2355	0.9908
	G	0.0233	not rej.	rej.	rej.	–	803.2818	0.9837	G	0.1064	not rej.	not rej.	not rej.	–	**813.9492**	0.9854
	W	0.0052	rej.	rej.	rej.	–	806.5562	0.9599	W	0.0149	rej.	rej.	rej.	–	834.7108	0.9698
	B	*0.0584*	–	–	–	–	***793.1358***	*0.9881*	B	*0.1261*	–	–	–	–	*816.8457*	0.9943
Setúbal	LN	0.042	rej.	–	–	rej.	643.6181	0.9849	LN	0.2064	not rej.	–	–	not rej.	992.277	0.993
	G	0.1266	not rej.	not rej.	not rej.	–	**638.5554**	0.9852	G	*0.2014*	*not rej.*	*not rej.*	*not rej.*	–	***989.7347***	*0.9927*
	W	0.0556	rej.	rej.	rej.	–	651.0384	0.9777	W	0.0069	not rej.	rej.	rej.	–	1006.3894	0.9693
	B	*0.1515*	–	–	–	–	*640.3451*	*0.9884*	B	0.0617	–	–	–	–	996.9953	0.9922
Tomar	LN	0.0001	rej.	–	–	rej.	960.1536	0.8192	LN	0.022	rej.	–	–	rej.	960.1207	0.972
	G	0.0101	rej.	rej.	rej.	–	914.3681	0.9735	G	0.0229	not rej.	rej.	rej.	–	951.9085	0.9934
	W	*0.0866*	*not rej.*	*not rej.*	*not rej.*	–	***879.4525***	*0.9035*	W	0.327	not rej.	not rej.	not rej.	–	949.0071	0.9911
	B	0.0628	–	–	–	–	881.3789	0.9886	B	*0.394*	–	–	–	–	***945.5549***	*0.9958*
Torres Vedras	LN	*0.4089*	*not rej.*	–	–	*not rej.*	***909.0043***	*0.9928*	LN	0.1257	not rej.	–	–	not rej.	**1023.587**	0.9941
	G	0.1393	rej.	rej.	rej.	–	913.9875	0.9877	G	*0.1797*	*not rej.*	*not rej.*	*not rej.*	–	*1024.8449*	*0.9908*
	W	0	rej.	rej.	rej.	–	953.1888	0.9245	W	0.0076	rej.	rej.	rej.	–	1054.3344	0.941
	B	0.1626	–	–	–	–	918.4306	0.9847	B	0.05	–	–	–	–	1034.7722	0.9914
Vila Real	LN	0.4941	not rej.	–	–	not rej.	847.587	0.9942	LN	0.4123	not rej.	–	–	not rej.	948.3044	0.9898
	G	*0.6311*	*not rej.*	*not rej.*	*not rej.*	–	***845.5931***	*0.9941*	G	0.6193	not rej.	not rej.	not rej.	–	**946.7769**	0.9674
	W	0.0646	rej.	rej.	rej.	–	866.3713	0.9671	W	0.048	rej.	rej.	rej.	–	979.43	0.9662
	B	0.4733	–	–	–	–	851.8828	0.9935	B	*0.6809*	–	–	–	–	*946.9656*	*0.9915*
Viseu	LN	*0.0463*	*not rej.*	–	–	*rej.*	***979.5536***	*0.9862*	LN	0.0124	not rej.	–	–	rej.	846.659	0.9836
	G	0.0054	rej.	rej.	rej.	–	984.7627	0.9778	G	0.0323	not rej.	not rej.	not rej.	–	**840.9689**	0.9906
	W	0	rej.	rej.	rej.	–	1015.1816	0.9207	W	0.0044	rej.	rej.	rej.	–	848.8726	0.9828
	B	0.0179	–	–	–	–	990.2789	0.9441	B	*0.06*	–	–	–	–	*841.3903*	*0.9894*

Tab. B.29: Goodness-of-fit tests' results for the marginal distributions fitted to the 6 stations: Santiago do Cacém, Setúbal, Tomar, Torres Vedras, Vila Real and Viseu, in Summer. The selected distribution is written in italic and the lowest AIC is in bold. LN - Lognormal; G - Gamma; W - Weibull; B - Burr

	Summer															
	$F(x)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2	$G(y)$	p-value χ^2	KS	AD	CvM	S-W	AIC	R^2
Santiago do Cacém	LN	0.0803	not rej.	–	–	rej.	516.9691	0.9841	LN	0.1963	not rej.	–	–	rej.	702.6556	0.9629
	G	*0.121*	*not rej.*	*not rej.*	*not rej.*	–	***513.4983***	*0.9915*	G	*0.28*	*not rej.*	*not rej.*	*not rej.*	–	*695.5295*	*0.9898*
	W	0.0107	rej.	rej.	rej.	–	523.5162	0.9782	W	0.0557	not rej.	rej.	rej.	–	693.3596	0.9842
	B	0.0973	–	–	–	–	515.0105	0.9914	B	0.2547	–	–	–	–	**689.2363**	0.9909
Setúbal	LN	0.0091	not rej.	–	–	not rej.	**621.2451**	0.9916	LN	0.1156	not rej.	–	–	not rej.	981.2871	0.9926
	G	0.0296	not rej.	not rej.	not rej.	–	621.8221	0.9725	G	*0.1305*	*not rej.*	*not rej.*	*not rej.*	–	***98.9831***	*0.9873*
	W	0.0007	rej.	rej.	rej.	–	655.007	0.9489	W	0.0015	not rej.	rej.	rej.	–	994.0749	0.9672
	B	*0.0819*	–	–	–	–	*625.3717*	*0.9942*	B	0.0288	–	–	–	–	987.4066	0.9836
Tomar	LN	0.0426	not rej.	–	–	rej.	810.3003	0.9467	LN	0.0179	rej.	–	–	rej.	919.1305	0.9489
	G	0.0666	not rej.	rej.	not rej.	–	813.0525	0.5548	G	0.0673	rej.	rej.	rej.	–	907.4869	0.9685
	W	0	rej.	rej.	rej.	–	903.9425	0.939	W	*0.3118*	*not rej.*	*not rej.*	*not rej.*	–	***886.283***	*0.9969*
	B	*0.0627*	–	–	–	–	***796.5495***	*0.8487*	B	0.263	–	–	–	–	889.4448	0.9936
Torres Vedras	LN	0.0002	not rej.	–	–	rej.	871.4171	0.9666	LN	*0.1768*	*not rej.*	–	–	*not rej.*	***1020.4658***	*0.9939*
	G	*0.0001*	*not rej.*	*rej.*	*not rej.*	–	***864.8148***	*0.9935*	G	0.0817	not rej.	rej.	rej.	–	1022.5363	0.9915
	W	0	rej.	rej.	rej.	–	876.4054	0.9537	W	0	rej.	rej.	rej.	–	1049.7967	0.9344
	B	0	–	–	–	–	869.6913	0.9923	B	0.0414	–	–	–	–	1034.4837	0.9744
Vila Real	LN	0.0852	rej.	–	–	rej.	754.9691	0.9471	LN	0.0035	rej.	–	–	rej.	839.4996	0.9502
	G	0.384	rej.	rej.	rej.	–	739.6054	0.9681	G	0.0207	rej.	rej.	rej.	–	829.2329	0.9899
	W	*0.9151*	*not rej.*	*not rej.*	*not rej.*	–	***721.2742***	*0.9967*	W	0.2746	not rej.	rej.	rej.	–	819.2402	0.9907
	B	0.8931	–	–	–	–	722.1724	0.9674	B	*0.4372*	–	–	–	–	***813.2236***	*0.9899*
Viseu	LN	0.0006	rej.	–	–	rej.	863.5636	0.9567	LN	0.0216	rej.	–	–	rej.	741.7038	0.9578
	G	0	rej.	rej.	rej.	–	880.7539	0.9398	G	0.0526	rej.	rej.	rej.	–	735.8487	0.8491
	W	0	rej.	rej.	rej.	–	950.6026	0.8374	W	0.0001	rej.	rej.	rej.	–	770.5351	0.9759
	B	*0.3252*	–	–	–	–	***839.4463***	*0.9849*	B	*0.1912*	–	–	–	–	***723.2099***	*0.9315*

Tab. B.30: Fitted distributions to the **observed wind** of 6 stations: Santiago do Cacém, Setúbal, Tomar, Torres Vedras, Vila Real and Viseu. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn			Winter			Spring			Summer		
	$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)
$\hat{k}_{SC}$	2.8851	(1.5177, 21.9135)	$\hat{\alpha}_{SC}$	5.0337	(4.2633, 6.0919)	$\hat{k}_{SC}$	2.298	(1.2837, 8.7221)	$\hat{\alpha}_{SC}$	55.5197	(46.6946, 68.3627)
$\hat{c}_{SC}$	3.086	(2.5928, 3.706)	$\hat{\beta}_{SC}$	1.0449	(0.8794, 1.2711)	$\hat{c}_{SC}$	5.8564	(4.914, 7.0853)	$\hat{\beta}_{SC}$	9.5123	(7.9957, 11.7247)
$\hat{\lambda}_{SC}$	0.139	(0.0571, 0.1867)	–	–	–	$\hat{\lambda}_{SC}$	0.1413	(0.1023, 0.1629)	–	–	–
$\hat{\alpha}_{S}$	7.8454	(6.5935, 0.5194)	$\hat{\alpha}_{S}$	4.9759	(4.2138, 6.0154)	$\hat{k}_{S}$	2.392	(1.2982, 9.2666)	$\hat{k}_{S}$	1.3961	(0.8586, 2.9248)
$\hat{\beta}_{S}$	2.1802	(1.8284, 2.6596)	$\hat{\beta}_{S}$	1.4068	(1.1776, 1.7113)	$\hat{c}_{S}$	6.0934	(5.1165, 7.3804)	$\hat{c}_{S}$	7.5132	(6.2956, 9.1841)
–	–	–	–	–	–	$\hat{\lambda}_{S}$	0.2015	(0.1468, 0.2311)	$\hat{\lambda}_{S}$	0.2176	(0.1865, 0.2393)
$\hat{\omega}_{T}$	2.529	(2.2955, 2.8281)	$\hat{\omega}_{T}$	2.4866	(2.2558, 2.7928)	$\hat{\omega}_{T}$	4.4189	(4.022, 4.9114)	$\hat{k}_{T}$	1.5565	(0.9487, 3.5933)
$\hat{\delta}_{T}$	5.318	(5.016, 5.615)	$\hat{\delta}_{T}$	5.5135	(5.1913, 5.8299)	$\hat{\delta}_{T}$	6.5479	(6.3517, 6.7382)	$\hat{c}_{T}$	7.7607	(6.4965, 9.4977)
–	–	–	–	–	–	–	–	–	$\hat{\lambda}_{T}$	0.1382	(0.117, 0.1516)
$\hat{\alpha}_{TV}$	6.4488	(5.45, 7.7746)	$\hat{\alpha}_{TV}$	7.5605	(6.4339, 9.1846)	$\hat{\mu}_{TV}$	1.8734	(1.8426, 1.9048)	$\hat{\alpha}_{TV}$	19.8309	(16.6146, 24.0177)
$\hat{\beta}_{TV}$	1.0528	(0.8806, 1.2814)	$\hat{\beta}_{TV}$	1.1962	(1.0094, 1.4661)	$\hat{\sigma}_{TV}$	0.243	(0.2201, 0.2651)	$\hat{\beta}_{TV}$	2.9204	(2.4474, 3.5449)
$\hat{\alpha}_{VR}$	5.4478	(4.6079, 6.5504)	$\hat{\mu}_{VR}$	1.3321	(1.2702, 1.393)	$\hat{\alpha}_{VR}$	12.4594	(10.5206, 15.0458)	$\hat{\omega}_{VR}$	5.1557	(4.6888, 5.7389)
$\hat{\beta}_{VR}$	1.3216	(1.1117, 1.6005)	$\hat{\sigma}_{VR}$	0.4802	(0.4357, 0.5207)	$\hat{\beta}_{VR}$	2.4946	(2.0941, 3.0137)	$\hat{\delta}_{VR}$	5.3493	(5.2163, 5.4871)
$\hat{\alpha}_{V}$	7.2782	(6.1393, 8.8576)	$\hat{\alpha}_{V}$	8.601	(7.288, 10.4743)	$\hat{\mu}_{V}$	1.8449	(1.8095, 1.8811)	$\hat{k}_{V}$	0.3866	(0.2585, 0.5744)
$\hat{\beta}_{V}$	1.0365	(0.868, 1.2648)	$\hat{\beta}_{V}$	1.2096	(1.0181, 1.478)	$\hat{\sigma}_{V}$	0.287	(0.2607, 0.3115)	$\hat{c}_{V}$	12.2891	(9.8334, 16.3119)
–	–	–	–	–	–	–	–	–	$\hat{\lambda}_{V}$	0.1956	(0.1848, 0.2055)

Tab. B.31: Fitted distributions to the **simulated wind** of of 6 stations: Santiago do Cacém, Setúbal, Tomar, Torres Vedras, Vila Real and Viseu. $\hat{\theta}_{MLE}$ is the parameters estimates obtained by MLE.

Autumn			Winter			Spring			Summer		
	$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)		$\hat{\theta}_{MLE}$	CI (95%)
$\hat{\alpha}_{SC}$	5.5457	(4.7319,6.7022)	$\hat{\alpha}_{SC}$	5.9182	(5.0461,7.15)	$\hat{k}_{SC}$	1.7845	(1.0689,4.5895)	$\hat{\alpha}_{SC}$	47.5812	(40.2341,58.2953)
$\hat{\beta}_{SC}$	0.9498	(0.8021,1.1569)	$\hat{\beta}_{SC}$	1.0035	(0.8454,1.2126)	$\hat{c}_{SC}$	8.0511	(6.7514,9.7629)	$\hat{\beta}_{SC}$	5.7955	(4.8694,7.1235)
–	–	–	–	–	–	$\hat{\lambda}_{SC}$	0.1119	(0.0941,0.1228)	–	–	–
$\hat{\alpha}_{S}$	6.7426	(5.7074,8.18)	$\hat{\mu}_{S}$	1.9398	(1.8893,1.991)	$\hat{\alpha}_{S}$	21.7899	(18.5018,26.4366)	$\hat{\alpha}_{S}$	24.1045	(20.3481,29.316)
$\hat{\beta}_{S}$	0.9143	(0.7646,1.1159)	$\hat{\sigma}_{S}$	0.3842	(0.3493,0.4199)	$\hat{\beta}_{S}$	2.2908	(1.9376,2.7892)	$\hat{\beta}_{S}$	2.4694	(2.0837,3.0105)
$\hat{\alpha}_{T}$	8.056	(6.7805,9.8343)	$\hat{\alpha}_{T}$	7.194	(6.0495,8.8881)	$\hat{k}_{T}$	4.3526	(1.7109,65.6576)	$\hat{\omega}_{T}$	6.9184	(6.3165,7.7267)
$\hat{\beta}_{T}$	1.0715	(0.8981,1.3182)	$\hat{\beta}_{T}$	0.9567	(0.7972,1.1901)	$\hat{c}_{T}$	7.332	(6.2344,8.768)	$\hat{\delta}_{T}$	10.5376	(10.3322,10.7368)
–	–	–	–	–	–	$\hat{\lambda}_{T}$	0.0838	(0.0502,0.0954)	–	–	–
$\hat{\alpha}_{TV}$	11.6712	(9.9382,14.2435)	$\hat{\alpha}_{TV}$	11.7825	(9.9231,14.3044)	$\hat{\alpha}_{TV}$	23.0983	(19.6645,27.8651)	$\hat{\mu}_{TV}$	2.2655	(2.2376,2.2934)
$\hat{\beta}_{TV}$	1.3119	(1.1126,1.6054)	$\hat{\beta}_{TV}$	1.3087	(1.0976,1.5984)	$\hat{\beta}_{TV}$	2.354	(1.9907,2.8537)	$\hat{\sigma}_{TV}$	0.2163	(0.1965,0.235)
$\hat{k}_{VR}$	1.5522	(0.9403,3.3703)	$\hat{\mu}_{VR}$	1.9571	(1.9147,2.0012)	$\hat{k}_{VR}$	1.5505	(0.9493,3.3292)	$\hat{k}_{VR}$	3.3683	(1.6679,41.8492)
$\hat{c}_{VR}$	4.856	(4.073,5.908)	$\hat{\sigma}_{VR}$	0.3397	(0.3072,0.3685)	$\hat{c}_{VR}$	8.2577	(6.9424,10.0088)	$\hat{c}_{VR}$	9.2399	(7.884,11.0484)
$\hat{\lambda}_{VR}$	0.1208	(0.0942,0.1405)	–	–	–	$\hat{\lambda}_{VR}$	0.1031	(0.0895,0.1126)	$\hat{\lambda}_{VR}$	0.0945	(0.0667,0.1048)
$\hat{\alpha}_{V}$	8.5741	(7.2868,10.3517)	$\hat{k}_{V}$	2.4009	(1.3245,10.6155)	$\hat{k}_{V}$	2.3697	(1.3487,8.2245)	$\hat{k}_{V}$	1.9686	(1.1554,5.3442)
$\hat{\beta}_{V}$	1.4289	(1.2097,1.7365)	$\hat{c}_{V}$	3.8043	(3.185,4.5991)	$\hat{c}_{V}$	7.2234	(6.1274,8.6451)	$\hat{c}_{V}$	0.5088	(7.9323,11.4037)
–	–	–	$\hat{\lambda}_{V}$	0.1322	(0.0764,0.1649)	$\hat{\lambda}_{V}$	0.1222	(0.0957,0.1365)	$\hat{\lambda}_{V}$	0.1304	(0.1114,0.1413)

Tab. B.32: Copulas selected according to the AIC to fit 6 stations: Santiago do Cacém, Setúbal, Tomar, Torres Vedras, Vila Real and Viseu. $\hat{\theta}_1$ and $\hat{\theta}_2$ are the copula parameters estimates obtained by MPLE.

Station	Season	Copula	$\hat{\theta}_{MPLE}$				Tail Dependence [1]		p-values		
			$\hat{\theta}_1$	CI (95%)	$\hat{\theta}_2$	CI (95%)	$\hat{\lambda}_L$	$\hat{\lambda}_U$	$S_n^{(B)}$	S_n	$S_n^{(K)}$
Santiago do Cacém	Autumn	C_α^{gu}	2.3077	(2.023, 2.5924)	–	–	0	0.6497	1	1	0.62
	Winter	C_ρ^{gaus}	0.8312	(0.7985, 0.8639)	–	–	0	0	0.988	1	0.35
	Spring	C_α^{gu}	1.6129	(1.435, 1.7907)	–	–	0	0.4631	0.982	1	0.67
	Summer	C_α^{gu}	1.4967	(1.3106, 1.6828)	–	–	0	0.411	0.814	1	0.4
Setúbal	Autumn	C_α^{gu}	1.9654	(1.749, 2.1818)	–	–	0	0.5771	0.742	0.984	0.84
	Winter	$C_{\rho\eta}^{t}$	0.7062	(0.6423, 0.77)	4.4525	(*)	0.3735	0.3735	0.283	0.0934	0.21
	Spring	$C_{\rho\eta}^{t}$	0.703	(0.6278, 0.7782)	4.523	(*)	0.3674	0.3674	0.767	0.9046	0.07
	Summer	C_α^{gu}	1.7593	(1.5437, 1.9749)	–	–	0	0.5171	0.751	1	0.73
Tomar	Autumn	$C_{\rho\eta}^{t}$	0.6732	(0.5789, 0.7676)	4.3712	(*)	0.3496	0.3496	0.834	0.3551	0.74
	Winter	C_α^{gu}	2.0018	(1.7915, 2.2121)	–	–	0	0.5862	0.987	1	0.85
	Spring	$C_{\rho\eta}^{t}$	0.6013	(0.5211, 0.6816)	6.248	(*)	0.2197	0.2197	0.613	0.1543	0.26
	Summer	$C_{\rho\eta}^{t}$	0.7204	(0.6526, 0.7883)	3.7061	(*)	0.4242	0.4242	0.964	0.7597	0.13
Torres Vedras	Autumn	C_α^{gu}	2.057	(1.7908, 2.3232)	–	–	0	0.5993	0.964	1	0.66
	Winter	C_α^{gu}	2.0007	(1.7449, 2.2565)	–	–	0	0.586	0.963	1	0.51
	Spring	$C_{\rho\eta}^{t}$	0.6827	(0.6157, 0.7497)	6.661	(*)	0.2652	0.2652	0.975	0.7288	0.54
	Summer	C_α^{f}	5.4875	(4.3723, 6.6027)	–	–	0	0	0.908	1	0.82
Vila Real	Autumn	C_ρ^{gaus}	0.7336	(0.6777, 0.7895)	–	–	0	0	0.972	1	0.55
	Winter	C_α^{gu}	2.2934	(1.9945, 2.5923)	–	–	0	0.6471	0.974	0.998	0.41
	Spring	C_α^{gu}	1.7519	(1.5423, 1.9615)	–	–	0	0.5146	0.993	1	0.73
	Summer	C_α^{sg}	1.4039	(1.2431, 1.5649)	–	–	0.3616	0	0.643	0.727	0.8
Viseu	Autumn	C_ρ^{gaus}	0.6869	(0.6239, 0.7499)	–	–	0	0	0.754	1	0.01
	Winter	C_ρ^{gaus}	0.7418	(0.689, 0.7947)	–	–	0	0	0.997	1	0.63
	Spring	C_α^{f}	2.1755	(1.3622, 2.9889)	–	–	0	0	0.954	1	0.51
	Summer	C_ρ^{gaus}	0.3496	(0.2427, 0.4564)	–	–	0	0	0.969	1	0.16

(*) At present the asymptotic variance cannot be fully estimated if η is not fixed, thus it is not possible to provide a confidence interval; see the package *copula* manual.
[1] the 0's are theoretical.

C | Scripts

In order to exemplify how to adapt a JAGS script to a WINBUGS script, we will use the Autumn season of Aveiro. The remaining R scripts for the fit of the marginals and the copulas can be found in `https://github.com/lidiamandre/copulas`.

C.1 JAGS

```
gumbel<-"model{
    for (i in 1:n){
        x[i]~dlnorm(mu1,prec1)
        y[i]~dlnorm(mu2,prec2)
    }
    prec1<-1/sigma1^2
    prec2<-1/sigma2^2
    alphac<-1/theta
    # Zero Tricks
    C<-10000000
    for(i in 1:n){
        zeros[i]~dpois(phi[i])
        phi[i]<--logL[i]+C
    # Distribution function of the Lognormal distribution
        u[i]<-plnorm(x[i],mu1,prec1)
        v[i]<-plnorm(y[i],mu2,prec2)
    # Log-likelihood of the marginals
        l1[i]<--log(x[i]*sigma1*sqrt(2*3.141593))-0.5*((log(x[i])-mu1)^2/sigma1^2)
        l2[i]<--log(y[i]*sigma2*sqrt(2*3.141593))-0.5*((log(y[i])-mu2)^2/sigma2^2)
    # Log-likelihood of the Gumbel-Hougaard copula
        a[i]<--log(u[i])
        b[i]<--log(v[i])
        w[i]<-a[i]^(alphac)+b[i]^(alphac)
        l3[i]<--log(u[i]*v[i])+(alphac-1)*log(a[i]*b[i])+
            log(w[i]^(2*(1/alphac)-2)+(alphac-1)*w[i]^((1/alphac)-2))-w[i]^(1/alphac)
    # Log-likelihood of the model
        logL[i]<-l1[i]+l2[i]+l3[i]}
    # Prior for the marginal parameters
        mu1~dnorm(0.0,0.0001)
        mu2~dnorm(0.0,0.0001)
        sigma1~dgamma(0.001,0.001)
        sigma2~dgamma(0.001,0.001)
        # Prior for the copula parameter
        theta~dbeta(0.5,0.5)}"
```

C.2 WINBUGS

```
model{
     for (i in 1:n){
x[i]~dlnorm(mu1,prec1)
y[i]~dlnorm(mu2,prec2)
     }
     # Zero Tricks
     C<-100000
     for(i in 1:n){
          zeros[i]<-0
          zeros[i]~dpois(phi[i])
          phi[i]<–loglik[i]+C
     # Distribution function of the Lognormal distribution
          u[i]<-phi((log(x[i])-mu1)*sqrt(prec1))
          v[i]<-phi((log(y[i])-mu2)*sqrt(prec2))
     # Log-likelihood of the marginals
          l.marginais[i]<-0.5*log(prec1)-0.5*log(2*3.141593)-log(x[i])
               -0.5*prec1*pow(log(x[i])-mu1,2)+0.5*log(prec2)-0.5*log(2*3.141593)
               -log(y[i])-0.5*prec2*pow(log(y[i])-mu2,2)
     # Log-likelihood of the Gumbel-Hougaard copula
          a[i]<–log(u[i])
          b[i]<–log(v[i])
          l.copula1[i]<- -pow(pow(a[i],alphac)+pow(b[i],alphac),1/alphac)
          l.copula2[i]<-log(pow(pow(a[i],alphac)+pow(b[i],alphac),1/alphac)+alphac-1)
          l.copula3[i]<-(1/alphac-2)*log(pow(a[i],alphac)+pow(b[i],alphac))
          l.copula4[i]<-(alphac-1)*log(a[i]*b[i])
          l.copula5[i]<–log(u[i]*v[i])
          copula[i]<-l.copula1[i]+l.copula2[i]+l.copula3[i]+l.copula4[i]+l.copula5[i]
     # Log-likelihood of the model
          loglik[i]<-l.marginais[i]+copula[i]
}
     # Prior for the marginal parameters
          mu1~norm(0.0,0.001)
          mu2~dnorm(0.0,0.001)
          prec1~dgamma(0.001,0.001)
          prec2~dgamma(0.001,0.001)
          sigma1<-1/sqrt(prec1)
          sigma2<-1/sqrt(prec2)
     # Prior for the copula parameter
          alphac~dunif(1.0,100)
}
```